



Determinants of Functional Decline in Community-Dwelling Older Adults

Robert Fieo

**PhD
The University of Edinburgh
April 2011**

Declaration

I, Robert Fieo, declare that this thesis has been composed by me and that this is my own work, except as specified. I further declare that this work has not been submitted for any other degree or professional qualification.

Date: _____

Signature: _____

Abstract

The overarching theme of this thesis is the prevention of progressive-type disability. Unlike catastrophic disability, progressive disability is gradual and more common in older adults. Because progressive disability can take years to develop, it is often conceptualized as a continuum, from less to more disabled. Disability prevention, by definition, is designed to identify people who are as yet nondisabled but at high risk for future functional decline by identifying an early functional state associated with increased risk of subsequent disability (Fried & Guralnik, 1997). This thesis sought to address two challenges associated with identifying an early functional state of disability. The first challenge relates to instrument calibration. Traditional instruments (based on self-report) used for assessing disability, scales of activities of daily living (ADL) and instrumental activities of daily living (IADL), were originally developed to describe levels of functional status in institutionalized older adults. Thus, these instruments poorly discriminate, as well as underestimate disability in the early stages of development. Poor discrimination refers to tasks or activities (i.e., scale items) that prove unresponsive to changes in a particular person's ability level. Performance measures on the other hand, such as walk time or grip strength, have proven to be quite responsive to early declines in functional status. Despite the popularity of performance measures used to assess health status in epidemiology or gerontological research, evidence suggests that they measure a somewhat different construct than self-reported activities of daily living. ADLs have a long history of use in the medical community, yet it has been proposed that the relative standing of ADLs, in relation to community-dwelling older adults, could be enhanced by improving construct validities that are at least equivalent to those of physical performance measures. Item response theory (IRT) methodology can be used to improve the structure of ADL scales so that they are more sensitive in detecting the early stages of functional decline within relatively high functioning older adults; a stage that has been shown to be more responsive to clinical interventions aimed at prevention of overt disability or frailty. IRT can improve ADL scales in multiple ways: by confirming an underlying uni-dimensional continuum of disability, establishing interval level measurement or item hierarchies, and increasing scale precision. As part of this thesis I conducted a systematic review of functional status scales, applied to community-dwelling older adults, which employed IRT procedures. The review was useful in that it draws attention to areas of functional assessment that can be improved upon, most notably, the topic of establishing interval level data and construct under-representation. Using data from the Cardiovascular Health Study, I was able to show that a common hierarchy of functional decline was observed for a diverse set of conditions and diseases that are prevalent among community-dwelling older adults. Such an indicator could be used to identify hierarchical declines relating to severity in diverse patient populations. Improvements in validity of functional status scales can also lead to the use of ADL-IADLs as potential determinates of disability, rather than simply acting as outcome measures of disability. Again using data from the Cardiovascular Health Study, I examined the

predictive power of IADL (mobility-type) items on later disability. Self reported difficulty in 2 or 3 of the most difficult IADL items increased the odds of being disabled eight years later by a factor of 3.5. The odds of being disabled fell to 1.9 for those reporting difficulty with one item. The second challenge of this thesis relates to defining determinants of functional decline that manifest themselves at the earliest stages of the disablement process. As previously stated physical performance measures have been shown to be sensitive to early stages of functional decline. However, can other measures, potentially spanning multiple domains, be used to identify those at high risk for future disability? In particular I was interested in whether psychosocial and cognitive variables could be used to detect changes in functional status at the preclinical stages of the disablement process. With regard to the Cardiovascular Health Study, I was able to show that, for subjects within the normal range of cognitive functioning, performance in the lowest quartile of the Digit Symbol Substitution Test resulted in a 2.2 increase in the odds of being disabled. Performance on this measure, as well as self-reported mobility noted above, could detect decrements in functional status as much as 8 years prior. With the use of the Lothian Birth Cohort sample I explicitly investigated the psychosocial domain. I found that the level of depressive symptoms increased the odds of being disabled by 56%. Again, these symptoms were assessed as much as eight years prior to self-reported disability. The general findings of this thesis indicate that refinements in ADL-IADL measures can aid in the detection of disability at the pre-clinical level, and that cognitive function and intra-individual factors play a pivotal role in speeding up or slowing down the disablement process.

Acknowledgments

I'd like to thank Professor Ian Deary and Professor John Starr who have set the bar in regards to efficiency and professionalism. Not to mention their prodigious efforts related to hammering and shaping me into a functioning academic.

I'd also like to thank Doctor Elizabeth Austin and Doctor Wendy Johnson for their added expertise and support.

Table of contents (Text)

Abstract	iii
Acknowledgments	v
Table of contents	vi
Introduction	1
Chapter 1: Activities of Daily Living	5
1.1.1 The utility of functional status measures.....	6
1.1.2 Self-reported functional status vs. performance measures.....	8
1.2 Determinants of functional status.....	10
1.2.1 Cognitive influences.....	10
1.2.2 Reserve constructs.....	23
1.2.3 Socioeconomic status domain.....	25
1.2.4 Psychosocial domain.....	27
1.2.5 Physical fitness domain.....	28
1.3. Summary.....	30
Chapter 2: Can item response theory be used to improve the measurement of disability?	34
2.1 Methods of scoring.....	34
2.1.1 Interval scaling.....	36
2.1.2 Estimating error for scales.....	37
2.1.3 Establishing a scale hierarchy—Guttman scaling.....	38
2.1.4 One Parameter Rasch models (1PLM).....	42
2.1.5 Rasch versus Likert-scoring models.....	44
2.1.6 Discriminatory power.....	46
2.1.7 Mokken model.....	46
2.1.8 Mokken scaling versus the Rasch model.....	47
2.1.9 The Two-parameter parametric model (2PLM).....	48
2.2 Summary.....	49
Chapter 3: Evaluating the sensitivity and accuracy of instruments used to assess functional status in community-dwelling older adults: a systematic review	52
3.1 Background.....	52
3.2 Methods.....	57
3.2.1 Data sources.....	57
3.2.2 Inclusion and exclusion criteria.....	57
3.2.3 Reliability.....	59
3.2.4 Validity.....	60
3.3 Results.....	62
3.3.1 Articles close to inclusion.....	62
3.3.2 Reliability.....	63
3.3.3 Construct validity.....	64
3.3.4 Confirming a hierarchy.....	66
3.3.5 Content Validity.....	66
3.4. Discussion.....	67
3.5. Conclusion.....	73

Chapter 4: A hierarchical decline of functional status across the most common chronic conditions/diseases experienced by older adults	79
4.1 Background	79
4.1.1 Diseases and chronic conditions - Arthritis	81
4.1.2 Diabetes	82
4.1.3 Chronic obstructive pulmonary disease/chronic obstructive lung disease	84
4.1.4 Coronary heart disease	85
4.1.5 Comorbidity	86
4.2 Method	88
4.2.1 Sample	88
4.2.2 Assessment of functional status	90
4.2.3 Statistical analyses	91
4.3 Results	93
4.4 Discussion	94
4.5 Conclusion	96
Chapter 5: Markers of preclinical disability in older adults	102
5.1 Introduction	102
5.1.1 Mobility reserve	102
5.1.2 The case for 'prehabilitation'	103
5.1.3 Muscle strength as a predictor	104
5.1.4 Chronic conditions	106
5.1.5 Cognition as a predictor	107
5.1.6 Association between cognition and muscle strength	109
5.1.7 Incorporating performance and self-report measures	110
5.2 Methods	113
5.2.1 Study population	113
5.2.2 Selected conditions	113
5.2.3 Predictor variables	114
5.2.4 Outcome measure	115
5.2.5 Covariates	115
5.2.6 Statistical Analysis	115
5.3 Results	117
5.4 Discussion	120
5.5 Conclusion	124
Chapter 6: The role of cognitive abilities and intra-individual factors in the disablement process	130
6.1 Introduction	130
6.1.1 Cognitive Reserve	131
6.1.2 Fluid Intelligence	131
6.1.3 Depression	132
6.2 Methods	135
6.2.1 Study population	135
6.2.2 Measures	136
6.2.3 Statistical analyses	137
6.3 Results	138
6.3.1 Direct effects on disability	138
6.3.2 Indirect effects on disability	140
6.4. Discussion	141
6.5 Summary	145
Chapter 7: Discussion	153
7.1 Limitations	159
7.2 Conclusion	161

Table of contents (Figures & Tables)

Figure 1.1	Lawton-Brody instrumental activities of daily living	32
Figure 1.2	The main pathway of disability proposed by Nagi	33
Table 2.1	Guttman analyses of seven ADL-IADL items	51
Figure 3.1	Flow diagram for manuscript selection	74
Table 3.1	Studies using item response theory to establish a hierarchy of decline in ADL-IADL Scales	75
Table 3.2	Scales establishing interval level data	76
Table 3.2	Scales establishing interval level data continued	77
Table 3.3	Studies establishing invariant item ordering	78
Figure 4.1	Increasing the sophistication of the disablement process	97
Table 4.1	Effect of arthritis compared with that of other chronic conditions on the development of mobility or activity of daily living (ADL) difficulty	98
Table 4.2	Impact of COPD on self-reported functional limitation	99
Table 4.3a	Mokken Scaling Procedure Applied to 'CHD Only' Group	100
Table 4.3b	Mokken Scaling Procedure Applied to 'Arthritis Only' Group	100
Table 4.3c	Mokken Scaling Procedure, Applied to 'Diabetes Only' Group	100
Table 4.3d	Mokken Scaling Procedure Applied to 'Arthritis & CHD' Group	101
Table 4.3e	Mokken Scaling Procedure Applied to 'Arthritis & Diabetes' Group	101
Table 4.3f	Mokken Scaling Procedure Applied to 'Arthritis & COPD' Group	101
Figure 5.1	Depiction of paper-based WAIS-III Digit Symbol Substitution Test	125
Figure 5.2	Flow diagram of subjects from the Cardiovascular Health Study	126
Table 5.1	Hierarchy for older adults endorsing chronic disease/conditions	127
Figure 5.3	Digit Symbol Substitution Test performance grouped by scaled scores	128
Figure 5.4	Digit Symbol Substitution Test performance grouped in quartiles	129
Figure 6.1	The inclusion of contextual and psychological variables	146
Figure 6.2	Sample item from Raven's Advanced Progressive Matrices	147
Figure 6.3	Late-life depression process	148
Figure 6.4	Flow diagram for Lothian Birth Cohort study participants	149
Table 6.1.	Descriptive Statistics for LBC21 at Baseline and Follow-up	150
Figure 6.5	LBC variables examined within the disablement model	151
Table 6.2	Correlations between disablement components	152
Appendix		
Table A1	Tasks of daily life adopted from the Health Interview Survey Supplement on Aging Questionnaire	198
Figure A1	Sample sheet of Townsend's Disability Scale	199

Introduction

Approximately 20% of people aged 70 years or older, and 50% of people aged 85 and over, report severe disability, i.e., difficulties in such basic activities of daily living as bathing, dressing, toileting, continence, feeding, and transferring from chair to bed (McGee et al., 1998). Disability prevalence rises strongly with age. The average global prevalence of moderate and severe disability ranges from 15% in adults aged 15–59 years, and 46% in adults aged 60 years and older (World Health Organization, 2004). However, in high-income countries like Britain a similar percentage of older adults reporting moderate to severe disability is not reached until the end of the 7th decade of life.

Recent estimates of life expectancy for British citizens indicate, at age 50, the average male may live to age 79, and for women this figure rises to 83 (Jagger, 2008). However, life expectancy is a rather crude indicator of population health. A relatively new metric for assessing population health considers whether older adults can maintain good health, thus evaluating the quality of life for those in the highest age ranges. This metric is often reported as disability-free life expectancy or healthy life expectancy. Today, more than 65 countries worldwide have estimates of health expectancy based on these surveys. The evidence suggests that both British men and women can expect to have as much as 10 years of disability-type symptoms prior to death (Jagger, 2008).

As noted above, typically, disability increases with age. Yet findings do suggest that some older adults avoid significant reductions in functional capacity, and that older age is not uniformly associated with declines in performance (Seeman et al., 1994). These individual differences have prompted researchers to define attributes that are associated with healthy aging. In addition to knowing which attributes appear to attenuate or speed up the development of disability, it is important to know when an individual enters a stage of gradual decline in functional status. This stage has come to be identified as pre-clinical disability. Pre-clinical disability is a diagnostic tool that has emerged from a life-course view of disability and successful aging, a view

which acknowledges that antecedents of late-life disability (which is more often a progressive disability) may occur earlier in life (Corvinsky et al., 2005). Pre-clinical disability serves to identify older adults at high-risk for future functional decline, but who currently are nondisabled. It is these pre-clinical individuals that can be targeted for preventive-type interventions. Programs designed to prevent functional decline in older adults show that participants with relatively good functional status or moderate frailty are those who benefit the most from ‘prehabilitation’ (Gill et al., 2000). Finally, risk reductions, in self-reported health or decreased disability and pain for example, have approximated 10% per year of intervention (Fries, Corey & McShane, 1997).

Identification of those at high risk for future disability can be aided by advancements in the instruments used to assess disability. Disability is most commonly assessed with Likert scales, which record the level of self-reported difficulty in everyday tasks. Traditional disability instruments were designed for low functioning, often institutionalised older adults. Thus, these disability instruments or functional status instruments need to be revised and calibrated to be more useful in community-dwelling samples of older adults. Modern test theory can be employed to improve the reliability and validity of functional status measures. Improving reliability and validity is crucial because one will need increased sensitivity and precision to detect meager traces of disability (i.e., early or pre-clinical stages of disability) in relatively high functioning samples.

If one uses a self-report Likert scale to measure level of disability, the scale or instrument needs to be reliable. Knowing the instrument's reliability provides information about the variance or error associated with the person's true score. The true score refers to the average score a person would receive if they were repeatedly administered parallel measures. Instrument reliability relating to disability can tell us whether observed changes are due to the, for example, the intervention or problems with the instrument. An unreliable disability instrument may therefore underestimate the size of the benefit obtained from an intervention.

Having stated the importance of reliability, it has been argued that validity is the most important criterion in judging the effectiveness of an instrument (Alagumalai & Curtis, 2005). In this thesis, I have attempted show that item response theory (IRT) can provide additional information, beyond classical methods, that can be used to enhance construct validity and content validity. Content validity is a non-statistical type of validity that involves “the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured” (Anastasi & Urbina, 1997, p. 114). For example, does an IQ questionnaire have items covering all areas of intelligence discussed in the scientific literature? Construct validity refers to whether a scale actually measures what the theory says it will measure. IRT improves content validity by converting self-reported difficulty on Likert-type scales (ordinal level) to interval level data. With such data, one can quantify more clearly the nature of ceiling effects, as well as identify construct under-representation, i.e., gaps along the construct being measured. Construct validity can also be enhanced with IRT measures, through the use of confirming a formal hierarchy of functional decline.

Construct validity for functional status scales can also be enhanced by formally confirming a hierarchy of decline; for example, by supporting or refuting the expectation that ‘Stepping over obstacles’ is a more challenging task than ‘Walking over a level surface’ (Chiu, Fritz, Light, & Velozo, 2006). Establishing a hierarchy of functional decline tells more than the typical simple summation of functional loss, and may have predictive value to the clinician monitoring older adults: if the sequence is accelerated or out of order it may indicate the need for interventions (Daltroy, Logigian, Iversen, & Liang, 1992). The IRT item hierarchy goes beyond ordering items by their mean level of difficulty; ordering items or tasks by group mean scores does not imply that this ordering also holds at the individual level. “Any set of items can be ordered by item mean scores, but whether such ordering also holds for individuals has to be ascertained by means of empirical research” (Ligtvoet et al., 2010, p. 579). Furthermore, scales revised through IRT procedures must adhere to assumptions relating to, for example, unidimensionality, local independence, monotonicity, and noninteresection of item response functions. It may

be, in order for a scale to meet these requirements, that some items must be removed from the scale. These actions will help to ensure that item ordering will be maintained at the individual level.

In addition to drawing attention to the utility of revised functional status instruments for use in community-dwelling older adults, this thesis will also seek to identify variables that exhibit strong associations with disability in late-life. Progressive disability can take years to develop, thus allowing room for appropriate preventive strategies. If progressive disability can be viewed as a continuum, then gerontology would benefit from identifying the pre-clinical interval along such a continuum. Identifying 'disability' in its earliest stages would allow for a larger window for successful attenuation, as well as improved targeting of interventions. Because I'm focused on the early stages of activity restriction it is essential that my investigation deals with older adults that are relatively high functioning. This thesis examines two cohorts of older adults still living within the community, the 1921 Lothian Birth Cohort (ages 79 to 87) and subjects from the Cardiovascular Health Study (ages 65 +). These cohorts were ideal for this thesis as very few subjects presented with disability at baseline assessment.

1

Activities of Daily Living

1.1 Defining functional status

In an interview (Beal & Katz, 2004) Dr. Sydney Katz recalled a conversation in the late 1950s with a colleague. Katz conveyed that, once he had a feeling for geriatric care, he got started on a project that required him to consult with George Badger. During the meeting, Badger stated, “You know, what you are talking about is something not generally talked about in medicine. I would call it function.” From here, Katz would go on to develop a scale (Katz, 1963) which is currently employed throughout the world. The scale, Activities of Daily Living (ADL) Assessment, was designed to define an individual’s ability to complete essential tasks of everyday functioning. The initial scale questioned whether a person was independent in the following: *bathing* (sponge bath, tub bath, or shower), *dressing* (gets clothes and dresses without any assistance except for tying shoes, *toileting* (goes to toilet room, uses toilet, arranges clothes, and returns without any assistance, *transferring* (moves in and out of bed and chair without assistance, and *continence* (controls bowel and bladder completely by self), *feeding* (feeds self without assistance). Soon thereafter, Lawton and Browdy (1969) would go on to develop the Instrumental Activities of Daily Living scale (see Figure 1.1, p. 32). This scale was intended to assess daily activities that were more complex than those assessed in the Katz ADL scale. The tasks assessed whether a person could use the telephone, go shopping, prepare a meal, do housekeeping, do laundry, the level or mode of transportation a person could manage, manage medication use, and lastly, the ability to handle personal finance. The scale was also more complex in its assessment of each task, i.e., it provided multiple response options that were meant to tease out the level of severity. These two scales mark the earliest attempts at assessing function in older adults, and since then, more than 75 instruments have been developed to measure functional status (McHorney & Cohen, 2000). These measures differ in the number of items,

type of rating scale, and item difficulty, but are typically still defined in terms of ADL-type or IADL-type. A majority of scales used today in epidemiological studies or clinical practice are likely to include both ADL and IADL items.

1.1.1 The utility of functional status measures

This thesis, like other studies preceding it, focuses on the transition from health to disability. Several models of transition have been proposed. The most well known include the International Classification of Functioning, Disability, and Health (ICF) framework of the WHO (World Health Organization, 2001) and Sudd Nahi's conceptual scheme of disability (Nagi, 1964). Today, both of these models are commonly referred to as the disablement process (see Figure 1.2, page 33), a term proposed by Verbrugge and Jette (1994). "The Disablement Process: (1) describes how chronic and acute conditions affect functioning in specific body systems, fundamental physical and mental actions, and activities of daily life, and (2) describes the personal and environmental factors that speed or slow disablement; namely, predisposing risk factors that propel dysfunction, interventions inserted to avoid, retard or reverse it, and exacerbators that hasten it" (Verbrugge and Jette, 1994, p.1). In this thesis I selected the Nagi model to provide the structure for investigations and interpretations into the disablement process. It has been argued that the Nagi model provides a better framework for the concept of a pathway or process (Guralnik & Ferrucci, 2009). This is particularly true for progressive disability as compared to catastrophic disability, with the latter being more common in younger individuals (Ferrucci et al., 1996). Progressive disability often takes years to develop which allows for sequential tracking through different stages. "The beauty of this framework [the Nagi model and its development over time] is that it allows for appropriate interventions at different points in the pathway" (Guralnik & Ferrucci, 2009, p.1171). The term "process" is used to reflect the dynamics of disablement, and has been formally defined as the trajectory of functional consequences over time and the factors that affect their direction, pace, and patterns of change (Verbrugge & Jette, 1994).

The main pathway of the disablement process is often described in terms of multiple domains, which progress from pathology, impairment, functional limitation, and ultimately disability (Nagi, 1965). Self-reported ADL-IADL measures often tap into the final domain of disability, which entails restrictions or the inability to perform socially defined roles and tasks expected of an individual within a sociocultural and physical environment. ADLs are typically presented as a Likert-type scale with respondents self-reporting on ordered levels of difficulty for a given task. Performance measures on the other hand typically assess the preceding third domain (functional limitations), and can include tasks such as bending over, or reaching overhead. Each of the broad domains (i.e., function or disability) reflects a critical step in the 'disablement process'. Verbrugge & Jette (1994) identify some of the most commonly applied dimensions of disability : (1) basic activities of daily living (ADL; often include basic mobility and personal care); (2) instrumental activities of daily living (IADL); (3) social roles, including occupation; (4) social activities, which might include attending church as well as socializing with friends and relatives; and (5) leisure activities, which might include physical recreation, reading, distant trips, and so on. As previously stated, each of the broad categories reflects a critical step in the 'disablement process', although in practice they are not completely distinct. For instance, some IADL scales (traditionally a disability factor) will include items that are more closely aligned to Nagi's functional limitations dimension (e.g., bending over or reaching over head).

Activities of daily living are essential in making a diagnosis of dementia (i.e., typically used to demonstrate decline and/or interference with work, social and interpersonal activities, or with the ability to function within society), relate to quality of life, and can reflect the level of economic burden required to maintain people with a disability. ADL measures have also been used to predict mortality (e.g., Inouye et al., 1998). Carey et al. (2004) used ADL, in conjunction with medical variables, to identify individuals at 36% risk of mortality in 2 years, which represents a 3-fold increase over expected mortality. Typically, for the elderly, ADL decline with age. For instance, McGee et al. (1998) found that less than 5% of men and woman in the under-70 age group were severely disabled compared with over one-

quarter of men, and over 40% of women in the over-85 group. Yet, such findings do suggest that some older adults avoid significant reductions in functional capacity. In an effort to identify and attenuate the observed health inequalities within the elderly population, researchers have sought to uncover the earliest stages of the disablement process. This has led to an increased use of performance measures (e.g., walk time) in epidemiological research which are thought to be more sensitive instruments than ADLs with regard to detecting the early stages of progressive-type disability. For instance, being compromised in one's ability to walk, lift, or climb (functional limitations) often leads to difficulty with employment and personal care, i.e., disability (McAuley et al., 2005). It is important to clarify the relationship between self-reported functional status and performance measures, as it is likely that they make independent contributions to the identification of health inequalities.

1.1.2 Self-reported functional status vs. performance measures

Activities of daily living (ADL) disability are commonly assessed through self-report, in which the individual is often given several Likert-type categories so as to ascertain severity. The impulse to question the validity of disability instruments because they rely on self-report, and to purport that performance based measures of ability are superior is not fully supported by the literature. This impulse relates to the subjective nature of self-reports, which tend to be influenced by mood, misjudgement of usual ability, or misinterpretation by the respondent (Louie & Ward, 2010). It is true that one of the most consistent determinants of self-reported decline in functional status is depressive symptoms. However, mood also seems to influence the outcomes on performance measures. Chou and Macfarlane (2009) found, rather surprisingly, that affective and personality characteristics (i.e., depressive symptoms and self-efficacy) had a larger influence on performance-based measures than on self-reports. This finding is not unique, as Cress et al. (1995) observed a negative relationship between affective status and performance measures. As Hoeymans (1997) suggests, these findings may relate to the fact that, whereas people's ability is dependent upon their current level of motivation, self-report allows for reports based on usual performance. Misjudgement of ability does occur in self-report measures but there is some evidence to suggest that misjudgement may often lead to an

underestimation of disablement. That is, people who do not have difficulty with physical performance measures have been identified as those people who also do not have problems with self-report measures (*low* false positive rate); at the same time there exists *high* rates of false negatives in which people self-report no problems with functional status items but perform poorly on physical performance tests (Branch et al., 2002; Sackett et al. 2000). Finally, the misinterpretation of items can be minimised by clinicians presenting the items in person, as well as the development of less ambiguous items.

Perhaps a less obvious, and often overlooked, disadvantage to the use of performance measures relates to the fact that they tend to reflect maximal ability in an artificial environment, which may not be an accurate reflection of performance in daily life (Coman & Richardson, 2006). One mechanism that is sure to underlie the disparity between two disability indicators, is adaptive functioning; performance-based measures do not reflect adaptations made in the person's everyday living situation (e.g., compensatory mechanisms such as handrails, or avoidance of difficult situations like walking outside in bad weather). Such adaptive responses to physical limitations may be influenced by factors such as personality and intelligence (Kempen et al., 1996).

Kempen et al. (1996) observed only moderate correlations between performance and self-report measures. Similarly, Chou & Macfarlane (2009) examined the association between self-reported and performance-based measures of lower extremity functioning and found a correlation of $r = .37$. They indicated that the performance and self-report measures were, at this level of association, related to different covariates. The covariates examined included sociodemographic variables, health indicators, cognitive function, depressive symptoms, and self-efficacy. Having said this, it should be noted that Kempen et al. indicated that their measures of performance and self-report did not refer to exactly the same activities. So, whereas it is true that personal and health characteristics influence self-report measures, strong associations can exist between self-report and performance measures. In fact, a recent systematic review of the relationship between self-report and performance

measures of functional status found that when instruments measure the same construct of the disablement process correlations tend to be quite large (.60 to .86) for instance a person's self-reported walking (either time or distance) was compared with observed walking, assessed with the same criteria (Coman & Richardson, 2006).

1.2 Determinants of functional status

Identifying the elements that underlie ADL measures can be considered the first step in providing practical interventions aimed at maintaining health with advancing age and the prevention of disability. Many studies have focused on ADLs as indices of disability, because ADLs represent the basic set of care needs essential to independent living (Saitoh, 2005). The determinants of ADL-IADL dysfunction are critical for the early identification of individuals at risk for functional disability and for improved patient care (Boyle & Cahn-Weiner, 2004). Family members and professional caregivers also rely on ADL reports to help compensate for areas of functional deficit (Weiner, 2006). Even modest advances in our understanding of the determinants of ADL may promote the use of appropriate compensatory strategies to prolong in-home living (Boyle & Chan-Weiner, 2004).

Throughout the 1990s researchers demonstrated that disability in the elderly is only partially accounted for by limitations in physical functioning, and that multiple factors are responsible for the disablement process (Kelly-Hayes et al, 1992; Seeman et al. 1995; Pennix et al., 1998). Louie & Ward (2010) assert that self-reported limitations in functional status represent a gestalt rather than an appraisal isolated from its context. That is, self-report ADL-IADLs cover a wide 'spotlight' of functional ability, beyond task difficulty within narrow or even artificial environments. In the remaining sections of this chapter, I will attempt to elucidate the varying components of the disablement process.

1.2.1 Cognitive influences

Much interest has been generated over the construct of general intelligence and how it can act as an index of health status. Scores from general intelligence tests have

been used to make causal inferences for human health and disease outcomes. Cognitive epidemiology has emerged as a relatively new sub-discipline within epidemiology and has focused primarily on outcomes associated with all-cause mortality. Deary & Batty (2006) conveyed that, given the well-established, or frequently replicated, links between total mortality and socio-economic status (SES), and the fact that intelligence is significantly associated with indicators of SES, there was support for investigating the association between cognition and specific diseases. It is now clear that the justification for such inquiries has been substantiated to some degree. For example, Australian Vietnam War veterans who presented with higher mental tests scores upon conscription were later shown to experience lower rates of overall mortality and motor accidents (O'Toole & Stankov, 1992). Furthermore, follow-up studies of the Scottish Mental Survey (1932) demonstrated an association between childhood IQ and survival to age 76 (Whalley & Deary, 2001). Most recently, the IQ relationship has been extended to disease specific outcomes, e.g., Hemmingsson et al. (2006).

Due to the multidimensional nature of IQ, further studies have sought to hone in on the subcomponents of cognition that support the IQ-mortality association. Deary & Der (2005), as well as Shipley et al. (2006) have shown that the IQ-mortality effect can be partially accounted for by simple and choice reaction time (RT), with Choice RT and RT variability having a greater influence than simple RT. This later point prompted Hall et al. (2009) to speculate that the relationship may have more to do with executive processes. Hall et al., followed 516 cognitively normal subjects who were free of serious medical conditions at baseline (mean age, 78.8) over a 10-year follow-up. Hall et al. assessed full scale IQ by averaging four Wechsler Adult Intelligence Scales-Revised (Wechsler, 1981), which included verbal comprehension, similarities (verbal reasoning), block design (visuospatial reasoning/motor ability), and digit symbol (executive function). They found that only digit symbol predicted mortality, after adjusting for demographics, education, and health behavior (includes smoking status, alcohol consumption, and exercise activities); for a 15-point decrease in digit symbol performance, subjects presented with ~ 29% increase in hazard of death. Digit symbol can draw on processing speed,

motor ability, but perhaps more centrally involves inhibition, working memory resources, and sustained attention to task (Lesso-Schlaggar et al., 2007). Finally, because the block design subtest, thought to also assess speed of information processing and motor abilities, was not associated with mortality, Hall et al. concluded that digit symbol is centrally related to executive processes.

Due to the fact that the reader of this thesis may not be familiar with fluid intelligence, I devote some necessary space to describing this construct. General intelligence (*g*) is a factor used to represent the variance shared by all measures of cognitive ability. As a mathematically defined entity with large explanatory power, the general cognitive factor has been exerting an influence in the psychological literature for over one hundred years (Blair, 2006; Spearman, 1904). Raymond Cattell's intelligence model (1963) includes two factors which are thought to support general intelligence, which include general fluid and general crystallized intelligence. Fluid functions have been shown in latent variable models to be essential aspects of general intelligence (Conway et al. 2003). Empirically, fluid reasoning (*Gf*) is the best predictor of performance on diverse tasks, to such a degree that *Gf* and general intelligence might *not* be psychometrically distinct at all (Carroll, 1993; Gray et al., 2004). Similarly, Carroll (2003) conveys that he is tempted to suggest that the reality of a Fluid Reasoning factor independent of *g* is at least questionable. Fluid intelligence generally refers to reasoning and novel, on-the-spot problem-solving ability and is thought to be related to metacognition (Gray, 2003), which rather simply stated refers to the ability to think about one's own cognitive processes. Using techniques based on classical test theory, fluid intelligence has been empirically defined as the latent trait extracted from a variety of reasoning-dominated tests, which Carroll (1993) originally conceptualised as a broad factor based on the common elements of tasks involving deductive and inductive reasoning, quantitative reasoning, cognitive flexibility, abstraction of common principles, the development of strategies, and manipulation of mental representations.

Using factor-analytic techniques, *Gf* has been empirically defined as the latent trait extracted from a variety of reasoning-dominated tests. Sub et al. (2002) describe four reasoning categories that might be subsumed under fluid abilities: *inductive reasoning*, which involves rules being abstracted from individual exemplars (typical inductive tasks are analogies, series completion, and matrix tasks); *deductive reasoning*, which refers to inferences that are true if the premises are true (typical tasks are syllogistic inferences, mathematical text problems, and surface development tasks); *construction*, which require the integration of given elements into a configuration with certain required features (typically assessed with the use of anagrams); and *planning tasks*, which require step-by-step action to produce an appropriate sequence to reach a given goal. Sub (2002) stipulates that these different reasoning categories are not separate mental abilities but, rather, are meant to clarify or systematize a rather large pool of reasoning tasks.

Despite Spearman (1904) arguing that *g* was what emerged from a large battery of tests (i.e., that it was not measured perfectly by any single test), psychometric research has attempted to create individual *g*-loaded tests. Tests that directly measure fluid cognitive functions have higher *g* loadings than do other cognitive measures (Colom et al. 2004). Such tests will involve complex cognitive operations such as reasoning and abstraction that is based on the complexity of the mental operations rather than specific knowledge in a given area (Arend et al., 2002). One such test is the Raven's Progressive Matrices (Raven, Court, & Raven 1977): much of the research that compares an individual's performance on the Raven's Progressive Matrices (RPM) to performance on other intelligence tests supports the notion that it is best conceptualised as a measure of general intelligence (Babcock, 2002). The items in RPM was designed specifically by John Raven (senior) to tap those 'eductive' (a neologism coined by Spearman) reasoning skills that Spearman had identified as the essence of *g*. RPM is based on abstract spatial relations, i.e., on nonverbal inductive reasoning. RPM has been used widely as a fluid intelligence test as well as a test that measures Spearman's *g* tasks (Jensen, 1998). RPM is considered by many as the 'gold standard' for inductive reasoning tasks.

Speed

In investigating the age-related differences in performance on the Raven's APM, Babcock (1994) found that processing speed and working memory accounted for a large proportion of the variance. With regard to processing speed, generalized slowing with age is thought to have a detrimental effect on cognition due to the inability to hold information on-line that is necessary for task completion (Bugg et al., 2006). It is argued that generalised slowing results in decreased efficiency, which then requires representations to be held longer in short-term memory buffers. This point, in part, relates to the assertion that the speed at which subjects sampled rules contributed to individual differences in age-related performance (Verguts & De Boeck, 2000). Slower processing also limits the amount of information that will simultaneously be available for processing, limiting the number of associations that can be created during study. Several longitudinal studies have implicated processing speed in relation to age-related cognitive decline (Finkle et al., 2004; Zimprich & Martin, 2002). Zimprich & Martin (2002), for instance, found that changes in fluid intelligence and changes in processing speed correlate at $r = .53$, and that approximately 28% of the variance was shared. Also, Salthouse (1999) indicated, after controlling for processing speed, that only 13% of age-related variance remained in reasoning ability (Raven's APM). Bugg et al. (2006) showed, using a similar fluid ability measure, Matrix Reasoning, (correlates at $r = .81$ with Raven's Standard Matrices) that 56% of the age-related reasoning variance remained after accounting for processing speed. A novel hypothesis is that greater speed (inspection time) is associated with higher intelligence because it represents more efficient perceptual representations which lead to lower working memory load (Gray et al., 2004). This latter point is supported by the finding that perceptual organization has a greater impact on item complexity than the amount of information presented (number of elements, number of rules, and number of rules applied to an element), with 53% of the variance accounted for by the former (Primi, 2000)

Memory load

With regard to working memory, individual differences in capacity determine the number of rules that can be managed on a particular problem (Vigneau et al., 2006). Carpenter, Just, & Shell (1990) performed an item analysis of Raven Progressive Matrices and found that an item's difficulty level (as evidenced by its error rate) was due to the number of rule (load) or tokens an item contained. Thus, Carpenter et al. speculated that the amount of information that can be maintained in working memory would be an important indicator of reasoning ability.

Goal management & coordination

Carpenter et al. (1990) also found that successful goal management (the generation of goals and sub-goals as well as coordination of those goals) was more likely among high-scoring subjects than average-scoring subjects. This is consistent with cognitive psychology literature which includes capabilities relating to the maintenance and manipulation of goals in its definition of fluid intelligence (Deary, 2000). In dealing with complex tasks related to an inductive geometric matrix (e.g., Raven's), executive-type abilities must be employed to cope with memory overload (Primi et al., 2000). That is, the central executive component of working memory assembles numerous elementary comparison processing loops (Klauer, 1990). Carpenter et al. (1990) seem to suggest that these loops were necessary for the analytical decomposition of the problem into smaller subgroups (Primi et al., 2000). These capabilities relate to the goal management component of *Gf* proposed by Carpenter et al.: the process of spawning sub-goals from goals, and then tracking the ensuing successful and unsuccessful pursuits of the sub-goals on the path to satisfying higher level goals of the global strategy (Primi et al., 2000). In addition to speed and working memory, Babcock (1994) also reported unique age-related variance attributed to other abilities; namely, coordination, which appears synonymous with Carpenter's coordination of goals, within the larger goal management ability.

The relationship between *Gf* and working memory may have more to do with the executive component and not capacity per se (Wiley et al., 2011), as discussed above in terms of goal-management. The coordination of several rules has more to do with

the central executive component of working memory. Perfetti et al. (2009) report that the strong executive process initiated during some fluid reasoning tasks is related to the different modulation of attention, with low-ability subjects (as compared to high-IQ subjects) relying more on executive components during moderate level reasoning tasks. For instance, Salthouse (1993) determined that the relationship between working memory and success on Raven's problems was constant across the test, despite the fact that rule complexity and thus capacity demands increase throughout the test. Similarly, Unsworth and Engle (2005) found that the relationship between working memory and Raven's was consistent across problems with varying degrees of rule complexity or rule tokens. Given these findings, and the fact that working memory (operation span) and Raven's have shown correlations as high as $r = .34$, Wiley et al. (2011) sought to clarify the ambiguity. Wiley et al. suggest that the relationship between working memory and Raven's has more to do with flexibility of attention and the ability to withstand distraction or interference, rather than capacity. This line of reasoning seems to be supported by Gray et al. (2003). Using fMRI, this study demonstrated that high-ability *Gf* subjects show greater activation in sub-regions of the prefrontal cortex that support attentional (executive) control. Gray et al. found that on high-interference trials related to lure stimuli, in both verbal and nonverbal working memory tasks (three-back), higher *Gf* performers were more accurate. Lures are non-target stimuli that require a higher demand for control to avoid misclassification. Here Gray et al. highlight the importance of attentional control in protecting goals, or other information held actively in mind, from interference. Within the context of performing Raven's problems, Wiley et al. suggest that when a new rule combination is needed that low working memory subjects may perseverate, and that high working memory subjects may present with greater flexibility in attention when shifting to a new rule is required.

Unsworth, Brewer, & Spillers (2009) have recently proposed a dual-component framework for interpreting individual differences in working memory capacity, which consists of controlled attention as well as retrieval abilities. The controlled attention components relate to *restrain*, *constrain*, and *sustain* (Poole and Kane, 2009). On the other hand, controlled retrieval abilities refer to information that can

not be maintained in the focus of attention (due to distraction and/or capacity constraints) has to be retrieved back into the focus via a cue-dependent search process (Unsworth, Spillers, & Brewer, 2009). It would appear that attentional control clearly relates to goal management and coordination abilities, with the retrieval abilities more closely aligned to capacity and memory load.

Abstraction

Babcock (1994) also reported unique age-related variance attributed to other abilities; namely, identification of rules (i.e., inducing abstract relations), application of such rules. According to Conway et al. (2002), the link between working memory (includes central executive component) and general fluid intelligence is the demand for active maintenance of information in the face of concurrent processing of information and/or attention shifts, and this is a demand for controlled attention. It is clear that controlled attention (Engle et al., 1999) contributes to the variance in fluid intelligence, but it is also clear that *Gf* is fundamentally more than controlled attention (Birney et al., 2006). Tasks that require controlled attention processes involve operations such as inhibition, divided attention, and set switching (Anum, 2006), but even with regard to these executive functions, Friedman et al. (2006) noted that only 57% of the variance in fluid intelligence was accounted for by executive abilities, namely, inhibiting, updating, and shifting.

The view that working memory is *not* synonymous with *Gf* is not uncommon among neuroscience or neuropsychology researchers (Kane et al., 2005). In fact, a recent meta-analysis concerning these two abilities concluded that that individual tests of working memory capacity and *Gf* (or reasoning) share only about 20% of their variance, on average (Ackerman et al., 2005). Garlick and Sennowski (2006) point out that, in addition to working memory and executive function, neurobiological models of fluid intelligence would be bolstered by the inclusion of abstraction, as all fluid intelligence problems involve abstraction. These authors also point out that abstraction has historically been recognized as the hallmark of intelligence (Synderman & Rothman, 1987). Both scaling studies and factor analyses show that tests that require reasoning, or reflect the products of past reasoning, fall in the

central parts of the scale, or load most highly on the general fluid ability (*Gf*) factors (Lohman, 2001). As previously stated, Sub (2002) described four reasoning categories that might be subsumed under fluid abilities: inductive reasoning, deductive reasoning, construction, and planning tasks. Of these categories, the most essential to understanding *Gf* is inductive reasoning abilities (Gustafsson, 1988). Inducing abstract relations is also relevant because this ability has been shown to account for the age-related variance in Raven's performance (Babcock (1994). Finally, according to Carpenter et al. (1990) of the process of problem solving, the processes that distinguish individuals are primarily the ability of goal management (the management of a large set of information in working memory) and the ability of rule inducing (finding the rule that correspond to the elements and figures within the matrix).

Application

Wang et al. (2008) add an additional ability factor; namely, deductive reasoning. Whereas induction means establishing rules, deduction would be the applying of rules (Shye, 1998). It is worth noting that Lohman and Lakin (2009) suggest that the distinction between these two forms of reasoning may exist more in the mind of the researcher developing a task than in the performance of examinees on that task, and that researchers have found that performance on deductive and inductive tests is strongly related (Wilhelm, 2005). Having said this, it still may be worth distinguishing between these two reasoning abilities; for one, because Babcock (1994) indicated that deduction (application), in addition to induction (identification), contributed unique age-related variance to Raven's performance. Sub et al. (2002) describes deduction as inferences that are true if the premises are true (typical tasks are syllogistic inferences, mathematical text problems, and surface development tasks). Arguably, the use of syllogisms in Raven's tasks is supported by the cognitive psychology view of mental model (Johnson-Laird, 2001), in that even deductive judgments applied to syllogisms without semantic content (e.g., 'All P are B. All B are C. So, all P are C') appear to make use of the visuospatial workspace or representation (Goel et al., 1997). Furthermore, Guyote and Sternberg (1981) reported significant correlations ($r = .35$ to $.60$) for adults between performance on

syllogism-solving problems and a spatial-abstract ability factor extracted from subtests of the Differential Aptitudes Test. Moreover, a direct relationship has been observed between deductive reasoning and the Raven's test: Bickersteth and Das (1981) revealed that performance on a syllogism-solving test was better for children with a high score as compared with a low score on the Raven Progressive Matrix.

Perceptual organization

Carpenter et al. (1990) also identified visual encoding or the creation of a mental representation as an important feature of matrix problem solving. In support of this finding, an investigation into the processes, strategies, and knowledge involved in problem-solving behavior for individual items (i.e., *Gf* construct representation) revealed that perceptual organization explained 53% of the variance in item complexity (Primi, 2000). Babcock (2002) notes that, in perceptually complex items, the likelihood of the formation of irrelevant groups of elements based on perceptual features is increased. Hence, such items impose demands on selective encoding and abstraction, because certain perceptual groupings must be ignored and others based on more abstract attributes considered (Babcock, 2002). The encoding process—the creation of internal mental representations of element attributes—is thought to be affected by perceptual organization involving the (1) the examining the elements of a figural matrices and the perceptual features of those elements, as well as (2) the perceptual grouping of the elements—the latter being based on Gestalt principles (Arendasy & Sommer, 2005). These forms of perceptual manipulation are also thought to influence the induction process related to the conflict between perceptual and conceptual groupings of elements within a matrix, which has a direct impact on item difficulty (Embretson, 2002).

This rather detailed section on *Gf* or Raven's might be considered superfluous in light of the fact that I do not and can not explicitly test for the above listed components. However, part of this thesis is intended to serve as a literature review concerning the determinants of functional decline, and in this way, a lengthy decision concerning *Gf* certainly seems warranted. This is particularly true if one considers

the prominent role that fluid reasoning has played in the literature pertaining to cognitive aging.

In part, epidemiologists were compelled to investigate the relationship between general intelligence and mortality because of the established patterning of mortality by socioeconomic status (SES), and the fact that intelligence is associated with SES. A similar line of reasoning appears applicable to the intelligence by disability association. As we have just mentioned, evidence supports a significant IQ and mortality relationship, and there also exists a relationship between disability and mortality. For instance, when attempting to construct the most accurate mortality index, researchers rely on disability measures. Thus, investigating the association between cognitive ability and disability appears warranted.

Cognitive epidemiology is interested in defining the individual differences that support optimal aging. Functional disability and cognitive impairment show strong and consistent associations. Dodge et al. (2005) studied a community-dwelling sample of elderly persons and found that cognitive impairment accounts for as much as 36% of incident disability (those who developed disability, but that were free of disability at baseline) in ADL tasks. In recent years noteworthy contributions have been made with regard to identifying the specific cognitive domains that support functional capacity. Impairment in instrumental activities of daily living is thought to reflect neurocognitive pathology associated with executive functions. Boyle & Cahn-Weiner (2004) demonstrated that executive dysfunction accounted for 28% of the variance in IADL. The association between executive functions and IADL remained significant even after adjusting for other cognitive functions (e.g., memory, attention and visuospatial skills). Additionally, Boyle et al (2002) reported that motor functioning in vascular dementia subjects accounted for only 14% of the variance in IADL. It has become quite clear that, “IADLs tend to decline earlier in the course of dementia than ADLs and scales which emphasize IADLs are most useful for outpatients with mild-to-moderate dementia. In contrast, scales which emphasize ADLs are most useful for inpatients or those with severe dementia” (Boyle et al., p. 110, 2003) However, healthy elderly populations present with a different profile, in

that, the more complex IADLs (e.g., preparing a meal) are maintained well after declines in more basic activities. For instance, Verbrugge, Yang & Juarez (2004), using a large community dwelling elderly sample (N=19,011), show that the frequency for disability in basic physical functions (e.g., reach, walk, or steps) is almost double the amount of disability observed for IADLs (e.g., meals or finances) and more than 4 times the frequency observed for self-care ADLs (e.g., bathe or dress). This pattern of results is very consistent with the findings in this thesis (i.e., chapter 5). Here it may be useful to note that Ng, Niti, Chiam, & Kua (2006) found that some commonly used IADL items (based on exploratory factor analysis) can be differentiated into physical IADLs and cognitive IADLs. Ng et al. show that a two factor model was a better fit than a one factor model and that the standardized regression coefficients of the cognitive IADL group was .35 vs. .20 for the physical IADL group. The former grouping included items assessing telephone use, taking medication, and managing money, while the latter grouping included items such as doing laundry, doing housework, and grocery shopping. This pattern of decline in community-dwelling older adults (i.e., some ADL items, but mostly physical IADLs items prior to cognitive IADLs) supports the view that dementia and age related cognitive change result from separate aetiologies. According to Meguro et al. (p. 565, 2001), “dementia is better conceptualized as an age-related (occurring within a specific age range) rather than as an aging-related disorder (caused by the aging process itself)”. Stated differently, dementia would not be conceptualized as one extreme in the continuum of natural aging.

In comparison to IADL there is less known about the underlying cognitive contributors to basic activities of daily living (ADLs). It is important to note, these scales share some overlap and the distinction between the two is often blurred in psychological literature. Both reflect behaviour in the service of everyday, simple tasks (e.g., tooth brushing) and extended activities (e.g., grooming), which require one to use objects and sequence multiple steps to achieve nested goals (Giovannetti, Libon, & Hart, 2002). Here nested goals reflect a situation in which the final goal is contained within a preceding set of goals, such that success in the final goal is dependent upon previous goals. The completion of everyday simple tasks may rely

more heavily on intact memory than executive ability. Jefferson et al. (2006) observed, among patients with vascular dementia, changes in memory over a one-year period were predictive of ADL changes. Also, Drachman et al. (1990) linked AD and memory functioning to ADL. Most noteworthy, Boyle et al. (2002) demonstrated, with the use of the Grooved Pegboard Test, that motor functioning accounted for 51% of the variance in ADLs .

In assessing the relationship between functional decline and cognition, we could postulate that a deteriorating brain is part of a body that is deteriorating generally, and thus body decline causes cognitive declines. However, Deary & Der (2005) relate that we need to be careful when inferring the direction of causality, because IQ of healthy children was shown to predict survival almost 70 years later just as well as IQ of middle age. Furthermore, low cognitive ability in childhood has been associated with an increased risk of dementia (Whalley et al., 2000). The early childhood effect reflects trait-like aspects of healthy brains that correlate with survival, or maybe low childhood IQ relates to earlier death partly because it is a reflection of a body with suboptimal physiological integrity (Deary & Der, 2005). It may be that cognitive ability early in life is an individual difference that buffers people from functional decline (i.e., ADL outcomes) through the accessibility of additional resources during exposure to life span stressors. It might not be unrealistic to assume that the mechanisms supporting the ADL/cognition relationship may be similar to those found in the cognition/longevity relationship. For instance, one explanation for the IQ/longevity relationship relates to the notion that people who present with higher IQ might interpret and respond more favourable to health messages, which might include information about fitness (Gottfredson & Deary, 2004). It is very plausible that this explanation might apply to the cognition/ADL relationship as well.

In both chapters five and six I explicitly investigate the relationship between cognition and ADLs. In chapter five, I test the influence of processing speed on the potential for endorsing self-reported disability. This is done over an eight-year time span. In chapter six I will examine how change in fluid reasoning ability impacts the

number of ADL tasks that a person indicates having difficulty with. This again will cover an eight-year time span.

1.2.2 Reserve constructs

The ability to avoid cognitive impairment, and perhaps functional disability, with advancing age may be attributable to one's level of cognitive reserve. The reserve hypothesis has been proposed to account for some of the individual differences observed in aging. Epidemiological evidence suggests that individuals who regularly participate in leisure activities, have a relatively high IQ, sufficient education, and high occupational attainment are less likely to develop Alzheimer's disease (e.g., Katzman, 1993). The concept of cognitive reserve posits that individual differences in how tasks are processed provide differential reserve against brain pathology or age-related changes (Stern et al., 2009). The reserve construct was initially proposed to account for the disparity between the presence of neuropathology and absence of behavioural dysfunction in some Alzheimer subjects. Perls (2004) provides a rather dramatic example of the reserve construct with a case series of 14 centenarians. All subjects were willing to undergo both neuropsychological and post-mortem neuropathological assessments. Four of the 14 subjects presented with a clinical dementia rating (CDR) of 0 and did not meet Braak and Braak criteria for AD. Upon autopsy these same individuals displayed no evidence of neuritic plaques on examination. Six subjects with CDR scores > 1 had neuropathological results consistent with their clinical presentation. In the remaining four cases, the subjects did not have CDR scores suggestive of dementia and yet they did meet 'Consortium to Establish a Registry for Alzheimer's disease' (CERAD) neuropathological criteria for possible or definite AD. An understanding of the individual differences of the 4 subjects that did not present with clinical symptoms or neuropathology is certainly warranted. However, cognitive reserve was meant to elucidate the differences in susceptibility once disruptions are present, and is not necessarily concerned with defining the various agents that cause pathology.

When researchers make reference to cognitive reserve it is often within the context of efficiency. Stern (2003) indicates that individuals with high reserve are probably using brain networks or cognitive paradigms that are more efficient or flexible. Thus these individuals may be less susceptible to various forms of disruption. This type of reserve is a normal process used by healthy people when confronted with an unusually difficult task. Yet these same individuals may also be better off when confronted with insults related to the normal aging process. For instance, white matter hyperintensities (WMH) are lesions that influence white matter integrity by causing cortical disconnections (Kanda et al., 2003). MRI scans reveal significant age-related differences in the frequency of WMH. Ylikoski et al. (1995) conservatively estimated that 65% of individuals over 75 years of age have white matter abnormalities. Increases in WMH predict poorer performance on tasks that tax speed of processing and executive control (Gunning-Dixon & Raz, 2003). Individuals operating from a modest level of reserve are in essence already taxed by lower levels of speed and control, so they are potentially nearer to the threshold of pathology. When increased demands arise as a result of age related pathology, someone who uses brain networks more efficiently, is more adept at calling into action alternative brain networks, or can adopt alternative strategies is more likely to maintain healthy functioning throughout the lifespan (Stern, 2003).

The reserve hypothesis identifies structural (passive) and functional (active) components that support the ability to compensate or adapt more easily to insults in normal or pathological aging. Passive reserve relates to anatomical features such as synaptic connectivity, neuronal count or the size of specific brain structures. In the broadest sense, passive reserve can be thought of as a model of brain capacity. Once this passive reserve capacity is depleted past its critical threshold, specific clinical or functional deficits emerge (Stern, 2006). Active reserve references one's ability to switch to alternative cognitive paradigms to overcome the effects of brain aging, or the ability to recruit compensatory neuronal structures to replace processing pathways damaged by aging (Staff, Murray, Deary, & Whalley, 2004). The active model is thought to account for the role of efficiency within the neuronal system. It is particularly relevant to researchers concerned with individual differences because it

asserts that the same amount of brain damage or pathology will have different effects on different people, even if passive factors (e.g., brain size) are held constant (Stern, 2006). For the most part, active reserve is not a reflection of gross anatomical differences between individuals (e.g., neuronal count), but rather it references an ability to process tasks in a manner that allows an individual to cope better with neuropathology associated with aging. From a reserve perspective, education and occupation attainment may reflect intellectual challenges experienced during life that accumulate reserve and allow cognitive function to be maintained in old age (Christensen, 2001; Gatz et al., 2001).

In chapter six of this thesis I examine the role that education and occupational status play in the disablement process. I not only examine the influence these variables have on disability, but also assess their influence at intermediate stages of disability, such as impairment and functional limitations. In chapter six I have a rather unique variable, which exhibits high face validity with regard to cognitive reserve, namely, an IQ score from subjects age 11. Given a person with a relatively high IQ at age 11, I would expect these subjects to report less disability at age 87 because, potentially, they have to exhibit larger levels of age-related pathology before the threshold for disability is reached, as compared to peers with much lower starting values. It should be noted that the conceptualization of reserve in the literature is made more complex by the fact that all of the reserve variables just mentioned, to greater or lesser degrees, influence or are influenced by the construct of socioeconomic status.

1.2.3 Socioeconomic status domain

Education, higher levels of psychometric intelligence and occupational level are commonly cited examples of how active reserve may serve to ameliorate cognitive declines associated with aging. Two of these three variables (i.e., education and occupation) have been studied extensively in association with socioeconomic status. The three most commonly employed indicators of socioeconomic status (SES) in contemporary industrialised societies are income, education, and occupation (Grundy & Holt, 2001). A large body of research literature exists endorsing the relationship between SES and mortality. In part, due to Bebbington's (1993) study with UK

elderly subjects, attention shifted to SES and disability. Bennington found that disability-free life expectancy (DFLE) between social classes was larger than differences in total life expectancy (TLE). DFLE combines mortality and morbidity by dividing TLE into various states to estimate the average years of remaining life a person can expect to live without disability (Mathews, Jagger, & Hanncock, 2005). Melzer et al. (2000) examined two groups (age 65 to 74 & 75 years and over, n = 10,377) and found the prevalence of disability was lower in social classes I & II (all those with professional & managerial occupations) than the remaining three classes. A classification of disabled included cognitive impairment, as measured by MMSE and the Automated Geriatric Examination Computer assisted Taxonomy. Melzer et al. found that the younger half of the first age group within classes III to V could expect nearly 3 years *less* of disability-free life expectancy than those in classes I & II.

Education has also been shown to exhibit a relationship with ADLs. Jagger et al. (2007) found that education level contributes to ADL at ages beyond 65, after adjustments for cognitive function. Also, van der Meer and Mackenbach (1998), using health indices related to ADLs found that those with the highest levels of education had general health perceptions that were more favourable than those with the lowest levels of education. Further, lower education resulted in less health improvements. Finally, Grundy & Glaser (2000) indicated that baseline severity of disability varied by age group, social class, educational qualifications and housing tenure for subjects (N=3,543) 55-69 years-old. At a 5-year follow-up, 36% of subjects had worse disability, with increased severity and new incidence of disability being associated with lower SES, baseline self-rated health status, age and gender. One explanation as to how education influences disability development pertains to adaptation. It has been proposed that education may have a positive affect on one's ability to adapt through modifying tasks or employing technical aides (Adamson, 2007; Jagger et al., 2007).

In the analyses presented in chapter five, I do not directly investigate the relationship between education and self-reported disability, but rather use education level as a covariate. However, in chapter six I explicitly test the independent contributions made by education and occupation and the odds of moving toward a state of disability. This is also done at stages of disablement that precede disability.

1.2.4 Psychosocial domain

Much of the variance in ADL declines that can not be attributed to SES is often explained in terms of psychosocial variables. Depression in particular seems to be the most influential psychiatric symptom experienced within community-dwelling elderly persons. Disability is significantly higher in depressed vs. non-depressed individuals and in some studies depression has proven to be even more influential than chronic medical illnesses (Nishiwaki et al., 2005; Wells et al., 1989). Ferring & Fillip (1995) report age effects for depressive symptoms; they found that, amongst the very old (75-92) positive affect was lower than that of the young-old individuals (65-75). In addition to depression's direct influence upon quality of life, it can also lead to increased rates of morbidity (Beekman et al., 1995). Apathy's influence over motivation is likely to interfere with ADLs via disruptions in initiative, persistence, planning and monitoring. Zawacki et al. (2002) studied a group of demented subjects (mean age=78.2) and found that apathy accounted for 27% of the variance in ADLs. Despite the study's use of demented subjects, it is likely that, to some degree, the findings can be generalized to a healthy elderly population because the researchers indicated that dementia severity was not significantly associated with ADLs. Zawacki et al. concluded that apathy may be an independent factor associated with functional decline, beyond cognitive abilities. Despite such findings, Penninx, Guralnik, & Ferrucci, (1998) point out that major depression is relatively rare among community-dwelling elderly and that only 1 to 2% are affected. However, these figures increase from 2 to 20% for those individuals who suffer from significant symptoms of depression but fall short of the DSM-IV criteria relating to major depression. Penninx et al. (1998) conducted a 4-year prospective cohort study that included 1,286 subjects 71-years and older. At baseline and 4 years post, the subjects performed a battery of physical fitness tests, and a depression scale. The depression

scale assessed depressive feelings and behaviours noted within a week of the evaluation date. After adjusting for health status and sociodemographic factors, the study found that as depressive symptoms increased, greater declines in physical performance over 4 years were noted (odds ratio for depressed mood vs. those without depressed mood was 1.55, [95% confidence interval 1.02, 2.34]).

Investigating depression may provide us with additional insight into unexplained variance in ADL and greater clarity in providing the most appropriate interventions. Kempen et al. (1999) studied a group of 574 low functioning older persons in a prospective cohort study that spanned a period of three years and concluded that preventing an increase in depressive symptoms may ameliorate further declines in physical functioning for individuals that already present with poor functioning. Additionally, Wang et al. (2002) studied 1,873 subjects aged 65 and older and found that depressed mood was associated with poor ADLs but also with increased rates of ADL decline over the follow-up period (subjects were followed biennially until dementia diagnosis or last visit before death). More recently, and perhaps the most significant findings relating to ADL decline and depression come from Lenze et al. (2005), who divided 756 community-dwelling subjects into three distinct groups (persistently depressed, temporarily depressed, and non-depressed). They assessed trajectory of depressive and functional disability and established longitudinal trends by following subjects for a period of three years (time points 1-4). The final model used binary logistic regression (outcome = no increase vs. increase) to communicate findings. Lenze et al. (2005) determined that 'persistently elevated' depressive subjects had a 5.3 greater odds ratio for increased disability than the group with 'temporarily elevated' group. Lenze et al. concluded that their findings support the hypothesis of long-term toxic effects of depression.

1.2.5 Physical fitness domain

When accounting for ADL declines, physical fitness variables reliably make significant contributions. The assessment of physical fitness variables for ADL should not be perceived as redundant, because as previously noted, Kelly-Hayes (1992) clearly differentiated between functional impairments that make the activity

impossible to perform and the actual competence with the activity. Furthermore, Rydwick, Frandin, & Akner, (2004), in addition to identifying cognitive training for attenuating ADL declines, proposed functional capacity and mobility interventions (the latter two representing distinct and separate domains). Rydwick et al. evaluated a customary motor task of walking 6 meters as a reference of lower extremity function. Dunlop, Hughes, & Manheim (1997) show that disability acquisition, within a community-dwelling sample (n=2,777), follows a hierarchical trajectory, with walking ability being compromised first. Generally, tasks requiring substantial lower-extremity abilities incur disability before those requiring mainly upper-extremity ones (Verbrugge, Yang, & Juarez, 2004). Declines in walking performance have been attributed to impairment of the oxygen delivery system and changes in muscle type with aging, which acts to increase relative effort (Brown and Holloszy 1991). Startzell et al. (2000) reported that elderly subjects require joint torque that exceeds available levels on customary motor techniques. With regard to torque, Hortobagyi et al. (2003) observed that older subjects (mean age 74) performed knee torque at levels that were 27% less than younger subjects (mean age 22). Measures of joint torque, as well as muscle strength are collected to derive relative effort. There appears to be a significant age effect for levels of relative effort (increases) which presses elderly subjects to carry out ADL tasks at high effort levels, and ultimately results in premature fatigue.

Relative effort can be expressed as the percentage of knee torque produced during a motor task (e.g., chair rise) in relation to the maximal knee torque produced at a similar knee joint position during a maximal-effort leg press (Alexander, Schultz, & Warwick, 2001). Hortobagyi et al. (2003) observed that 30% of the variance in relative effort was accounted for by muscle coactivation, specifically the amount of hamstring activity relative to the vastus lateralis activity was 1.6 times greater in older than younger subjects. This increased coactivity around the knee joint has been interpreted as a compensatory strategy. Such findings prompted the use of a six meter walk test so as to assess the influence of physical effort on ADL.

Physical fitness for upper extremities is also important for the successful completion of ADL. “To maintain independent living, older people must be able to perform basic tasks of daily living, such as hand gripping, lifting and transferring. Many of these tasks require adequate upper-body muscle strength to complete” (Forrest, Zmuda, & Cauley, 2007, p. 140). Some research indicates that muscle strength, over aerobic fitness, is more essential for completing such tasks (Phillips & Haskell, 1995). The justification for the use of grip strength as a predictor variable is more appealing when we consider that it correlates well with total body strength (Rantanen et al., 2003). For elderly samples, it has shown correlation coefficients with knee extensors of .47 to .51, trunk extensors .33 to .56, trunk flexors of .37 to .58. (Brach & VanSwearingen, 2002). Giampoli et al. (1999) assessed an elderly population 4 years after baseline assessment revealed ‘no disability’. They divided the sample (n=422, age 71-91 [controlled for age]) into quartiles based on grip strength performance (1st quartile representing weak performance). They found, at the end of the four year period, that 48% of individuals in the highest quartile exhibited incidence of disability, in contrast to 25% of individuals in the lowest quartile. Giampaoli et al. also quantified group performance as follows: a difference in strength of 10 kPa was associated with 17% risk of developing incident disability over the next 4 years. Finally, Forrest, Zmuda and Cauley (2007) tested a cohort of elderly woman (n=7,969, mean age 71.6) for changes in grip strength, as well as changes in correlates. For the age range of 75-79, subjects decline in grip strength over the ten year period was 24.3%. They also reported that difficulties in ≥ 2 functional activities resulted in a 30% increase in the odds of recording lower grip strength (one standard unit) .

1.3. Summary

It is important to note that Kovar and Lawton (1994) indicated that ADLs were initially formulated to assess the functional status of chronically ill or institutionalized elderly. Thus, they may be ineffective when evaluating community-dwelling populations in which researchers have to identify very low levels of disability. Currently, traditional ADL-IADL scales are largely ineffective at detecting early stages of disability when applied to community-dwelling samples. This will

result in large ceiling effects, with a large proportion of subjects being ‘unmeasured’. Gerontology has only just begun to develop functional status instruments that are sensitive enough to detect early declines and precise enough to detect change over time in these high functioning older adults.

Improvements in understanding the progression and severity of the disablement process may be enhanced by establishing item hierarchies. Investigating the hierarchy of disability in ADLs, may be useful in determining the most appropriate prevention strategies to delay the onset of disability. Developments in this area have been hampered somewhat by a lack of consensus on the nature of the disability construct. For one, a number of researchers have endorsed a multidimensional model of ADLs (e.g., Fitzgerald et al., 1993). And yet, others would reject such an approach in favour of a unidimensional model (e.g., Katz-based assumption—Katz et al., 1963). Proponents of a multidimensional model often endorse a ‘physical IADL’ and a ‘cognitive IADL’, and others still would include a third dimension (Stump et al., 1997) of basic self-care. Recently, Ng et al. (2006) found that Exploratory Factor Analysis of the Lawton and Brody IADL scale supported a two-dimensional (cognitive and physical) structure. A three dimensional model of ADLs has been proposed by Wolinsky and Johnson (1991); Wolinsky and Johnson uncovered a three dimensional disability structure after performing factor analytic techniques on ADL/IADL items. The trend of ADLs moving toward a multidimensional model may accelerate as it becomes more essential to service the needs of an increasing community-dwelling elderly population. Perhaps the tasks assessed will change in complexity, or maybe ADLs will evolve to emphasize and measure the level of efficiency with which a person completes a task.

In the next chapter I will discuss the theory behind attempts to improve instruments (ADL-IADL scales) used in the assessment and tracking of disability. I provide a basic introduction into item response theory (IRT), which includes the assumptions that are common to various IRT models. There exists a multitude of IRT models. The chapter will discuss the most commonly used models, which include both parametric and nonparametric types.

Figure 1.1

LAWTON-BRODY INSTRUMENTAL ACTIVITIES OF DAILY LIVING SCALE

ABILITY TO USE TELEPHONE

- | | |
|---|----------------------------|
| 1. Operates telephone on own initiative | <input type="checkbox"/> 1 |
| 2. Dials a few well-known numbers | <input type="checkbox"/> 1 |
| 3. Answers telephone but does not dial | <input type="checkbox"/> 1 |
| 4. Does not use telephone at all | <input type="checkbox"/> 0 |

SHOPPING

- | | |
|---|----------------------------|
| 1. Takes care of all shopping needs independently | <input type="checkbox"/> 1 |
| 2. Shops independently for small purchases | <input type="checkbox"/> 0 |
| 3. Needs to be accompanied on any shopping trip | <input type="checkbox"/> 0 |
| 4. Completely unable to shop | <input type="checkbox"/> 0 |

FOOD PREPARATION

- | | |
|--|----------------------------|
| 1. Plans, prepares, and serves adequate meals independently | <input type="checkbox"/> 1 |
| 2. Prepares adequate meals if supplied with ingredients or heats meals on wheels | <input type="checkbox"/> 0 |
| 3. Prepares meals but does not maintain adequate diet | <input type="checkbox"/> 0 |
| 4. Needs to have meals prepared and served | <input type="checkbox"/> 0 |

HOUSEKEEPING

- | | |
|---|----------------------------|
| 1. Maintains house alone or with occasional assistance | <input type="checkbox"/> 1 |
| 2. Performs light daily tasks such as dishwashing or bed making | <input type="checkbox"/> 1 |
| 3. Performs light daily tasks but cannot maintain acceptable level of cleanliness | <input type="checkbox"/> 1 |
| 4. Needs help with all house maintenance tasks | <input type="checkbox"/> 1 |
| 5. Does not participate in any housekeeping tasks | <input type="checkbox"/> 0 |

E. LAUNDRY

- | | |
|---------------------------------------|----------------------------|
| 1. Does personal laundry completely | <input type="checkbox"/> 1 |
| 2. Launders small items by self | <input type="checkbox"/> 1 |
| 3. All laundry must be done by others | <input type="checkbox"/> 0 |

MODE OF TRANSPORTATION

- | | |
|---|----------------------------|
| 1. Travels independently on public transport or drives own car | <input type="checkbox"/> 1 |
| 2. Arranges own travel via taxi, but does not use other modes of transportation | <input type="checkbox"/> 1 |
| 3. Travels on public transportation when accompanied by another | <input type="checkbox"/> 1 |
| 4. Travels limited to full assistance by another | <input type="checkbox"/> 0 |
| 5. Does not travel at all | <input type="checkbox"/> 0 |

RESPONSABILITY FOR MEDICATIONS

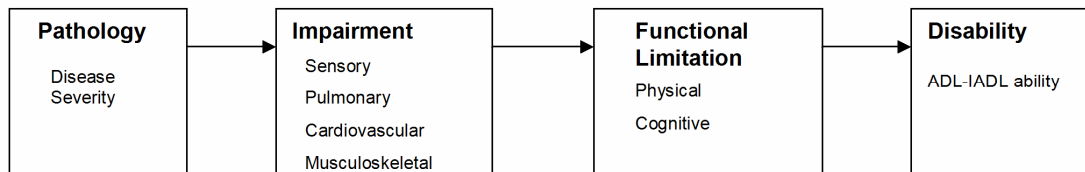
- | | |
|---|----------------------------|
| 1. Is able to take medications in correct dosages at correct time | <input type="checkbox"/> 1 |
| 2. Takes medications if they are prepared in advance in correct dosages | <input type="checkbox"/> 0 |
| 3. Is not capable of dispensing own medications | <input type="checkbox"/> 0 |

ABILITY TO HANDLE FINANCE

- | | |
|--|----------------------------|
| 1. Manages financials matters independently (e.g., paying bills, go to bank) | <input type="checkbox"/> 1 |
| 2. Manages day-to-day purchases but needs help with banking, major transactions. | <input type="checkbox"/> 1 |
| 3. Incapable of handling money | <input type="checkbox"/> 0 |

Figure 1.2

The main pathway of disability proposed by Nagi



The pathology stage refers to the diagnosis of disease, injury, congenital/developmental condition. The impairment stage describes dysfunctions and structural abnormalities in specific body systems. The functional limitations stage refers to restrictions in basic physical and mental actions. The final stage indicates one's level of difficulty in doing activities of daily life, such as personal care or household management.

2

Can item response theory be used to improve the measurement of disability?

Experience is continuous. But the moment we notice an experience, it becomes discrete. We sense the fragrance of flowers. The sensation is continuous. But when we distinguish between flowers—with and without fragrance; strong from weak fragrance, fragrance we like, don't mind, or dislike, then our observations become discrete. As we notice and remember particulars, we begin the counting that can be measurement...As soon as we start counting, we have decided on a useful identity, namely that, at least for us, the objects we count are sufficiently identical to be infinitely exchangeable (Wright & Stone, 2004).

2.1 Methods of scoring

Within the context of healthy aging, the 'useful identity' might be disability—more or less disability. Evaluating the degree of difficulty in performing activities of daily living (ADL) is a common way to measure physical disability. As indicated in the previous chapter, individuals are assessed on how much difficulty they have in performing activities of daily living such as walking around in their home, getting up from a chair, or walking two blocks. Standardized measurement of functional status began in the late 1950s. ADL scales vary in task complexity as well as the degree of difficulty for each task, i.e., number of options. A typical ADL scale assess task difficulty with a numerical label (e.g., 0 for 'no difficulty'; 1 for 'some difficulty'; 2 for 'unable to perform'). The most common scoring method for ADLs is Likert summative scoring (Likert, 1933) (McHorney, Haley, & Ware, 1997). This traditional aggregate method of scoring computes a raw total score by summing responses to individual items. There exists some controversy over whether Likert-based functional status can be combined to provide a relatively continuous and scalable variable at the 'interval' level (Hurst, 1994). Despite the fact that this scoring method is quite common in health research, there are well-established problems with raw scale scores that make them difficult to meaningfully interpret (Reise & Henson, 2003). It has been strongly argued that when ordinal-type items, which vary in difficulty, are summed the properties of true measurement are

compromised (Doble & Fisher, 1998). The Likert summative scoring method is based on classical test theory (CTT), which uses raw scores to compute statistics relating to means, variances, reliability coefficients, total scores and errors of measurement for an instrument as a whole (Wilson, Allen, & Corser, 2006). Modern test theory, which includes item response theory (IRT), can improve upon some of the statistics just mentioned, such as errors of measurement; IRT, unlike CTT, provides errors of measurement at the individual level, rather than an error score that is an average of all respondents. Standard errors of measurement are important because they inform the researcher about instrument precision. IRT is also useful in that it provides new information, above and beyond that which has traditionally been provided by CTT, thus increasing an instrument's interpretive power. For example, many IRT models can be used to convert ordinal Likert-type responses into interval level data.

The total score method assumes that each item or symptom on the scale represents an equal level of severity, which is almost never true (Gibbons, Clark, & Davis 1985). Likert scales may maintain an ordinal relationship within a single item, and yet sets of Likert items are not necessarily ordinal with respect to each other. Summative scoring models retain 'ordinal' properties, which means, applying mathematical procedures is not sound practice. For instance, average scores obtained from unrevised ordinal data are non-linear and thereby lack a fundamental property assumed in all measures, i.e., that the measures increase along a linear scale (Wright & Linacre, 1989). Furthermore, "The development of measurement instruments with themes relating to behavioural science is inherently more difficult [in relation to the physical sciences] because the properties being measured do not lend themselves equally easy to direct observation with a method that is simply and universally accepted" (Sijtsma & Molenaar, 2002, p.17), i.e., measuring length in the metric system. Psychometricians may begin with ordinal instruments (e.g. ADLs or personality rating scales), but they often seek to establish a universally accepted continuum for rating a construct, such as personality of disability, so that these instruments behave much like an inch-mark on a ruler (Sijtsma & Molenaar, 2002).

2.1.1 Interval scaling

The simple algebraic sum of item scores has been criticized, in part, because of its potential for misinference (McHorney et al., 1997); Thorndike (1904) identified problems inherent in using measurements of this type, such as the inequality of the units counted and the non-linearity of ‘raw scores’ (Sheehan et al., 2000). The potential for misinference is rooted in the inequality of scale intervals (Vittengl, White, McGovern, & Morton, 2006). Health status scores do not have ratio characteristics and thus one should not conclude that a patient with a score of 50 is twice as healthy as a patient with a score of 25 (Stucki, Daltroy, Katz, Johannesson, & Liang, 1996). A functional ability scale that assigns a numerical label to the task of ‘climbing stairs’ (e.g., 0 for ‘no difficulty’; 1 for ‘some difficulty’; 2 for ‘unable to perform’) does not make these numerals become equally distanced measures (Wright & Mok, 2004). Similarly, there is no rationale to assume equal meaning of changing from one item option to the next, say for example, accounting for the distance between ‘some difficulty’ and ‘no difficulty’ (Lundgren-Nilsson et al., 2005). More importantly, it is wrong to assume equal meaning of these steps for different task items. For instance, a change from ‘limited a lot’ to ‘limited a little’ for ‘going up and down stairs’ may not have the same significance as changing from limited a little to not limited in bathing. With unequal spacing between categories, it becomes extremely difficult to make inferences concerning health status. Number labels or raw counts may give the impression that they are interval (or ratio) measures of experience, however, this is not the case (Wright & Mok, 2004). Due to the fact that health status instruments are summed scores and typically include zero it has been common to treat them as continuous variables with ratio or interval characteristics but, in this case, the definition of a zero point is arbitrary and instrument dependent (Stucki et al., 1996).

In contrast the traditional Likert (Likert, 1932) summative scoring methods, item response theory (IRT) models meet the conceptual requirements of order and additivity (McHorney et al., 1997). This is because IRT is capable of transforming non-linear ordinal scores into linear measures. “With the priority placed on establishing interval units of measure, complementary tools for understanding the

nature of scale's meaning and, more importantly, providing a substantive context within which an individual's score on a scale may be interpreted" (Haley et al., 2004, p.52). IRT models specify the correlation between the observed response behaviour of a respondent and the respondent's localization on a latent trait (Farin et al., 2007). IRT makes use of item characteristic curves (ICCs) or Item Response Functions (IRFs) to demonstrate the degree to which response behaviour of each item are related to the unobservable latent trait (e.g., disability). The ICCs are expressed in graphical format with trace lines reflecting the probability of responding positively to an item with a person's position along a given latent trait.

2.1.2 Estimating error for scales

In addition to the concept of interval measure, error measurement is very relevant when revising summative Likert scales to increase interpretive power. Error measurement is relevant because it is treated differently in summative scoring; IRT is able to estimate measurement error more accurately or more precisely than Likert-type models. IRT parameters (e.g., difficulty & discriminatory power) are estimated to minimize error (Embretson & Reise, 2002).

Most investigations relating to ADL scale development and refinement emphasise aspects of reliability, i.e., overall measurement precision (McHorney, 2002). Reliability refers to the percent of observed score variance that is due to true score variance, where true score is defined as the average score an individual would receive if they were repeatedly administered parallel measures. Internal consistency, along with reproducibility, are measurement properties used to estimate scale reliability that assesses the extent to which the instrument is free from random error or noise (Alagumalai & Curtis, 2005). There is always some amount of random error which may push an individual's observed score higher or lower than the true score. The most commonly used approach to estimating internal consistency is Cronbach's alpha (Cronbach, 1951), which essentially estimates the average level of agreement of all the possible ways of performing split-half tests. Measuring internal reliability is a very important property of any clinical-based outcome measure. Reliability allows the researcher to confirm that observed changes are due to the intervention and not to

inaccuracies in the measurement tool. An unreliable measure may therefore underestimate the size of benefit obtained from an intervention (Fitzpatrick, Davey, Buxton, & Jones, 1998). As mentioned above, reproducibility is another component of reliability. Reproducibility, stated simply, is the evaluation of whether an instrument yields the same results on repeated applications, which is assessed by test—retest reliability. The degree of agreement is examined between scores at the first assessment and when reassessed.

Classical assessments of test reliability provide an overall evaluation of the precision of measurement, whereas IRT methods allow for the evaluation of measurement precision as a function of person ability level (Meijer & Sijtsma, 1990). A key limitation of summed-score methods is that the reliability and standard error of measurement (SEM) is constant for all subjects, irrespective of individual observed score level; it is assumed that the measure is equally precise for all persons, regardless of their ability or standing on the ability construct (Reise & Henson, 2003). IRT methods will provide different SEMs depending on where they are located on the latent construct (Bond & Fox, 2001). This is important for ADL analyses because it is often more valuable to examine a measure's information/precision across the entire construct than it is to know a single reliability coefficient.

2.1.3 Establishing a scale hierarchy—Guttman scaling

In part, to compensate for shortcomings in summing ordinal rated items, which may impede interpretation, the earliest developers of ADL scales adopted Guttman scaling procedures (Guttman, 1944). Guttman scaling has been used to determine if ADL items can be hierarchically arranged (Katz et al., 1963). With the Guttman model, scale investigators seek to determine whether the characteristics of a set of items fulfil two special conditions that define an acceptable Guttman Scale: unidimensionality and cumulateness (Kempen Myers & Powell, 1995). The model places the items in a 'continuum' depending on their level of difficulty. Guttman scaling groups subjects together by their total score (summed) in an effort to assess whether consistencies in person response patterns supports item ordering. Table 2.1

on page 51 depicts the Guttman analyses for seven functional status items, with a coefficient of reproducibility of 0.89 and a coefficient of scalability equal to 0.62.

As can be seen from the far right column, approximately 39% of subjects have none of the disabilities included in the analysis. Of those respondents who have one single disability (17.3%), dependence in cleaning is the most common (61.6% of those with one disability). On the other hand, the most infrequent disability is the one related to bed-making; dependence in this activity tends to imply dependence in all other activities. If items were perfectly ordered, in a cumulative fashion, the cells below the diagonal would all be 0, and the cell above the diagonal would be 100. In table 1, the coefficient of reproducibility (C of R) is slightly below the minimum requirement of 0.90. The deviations from the desired 'diagonal structure' acts to lower the C of R. C of R is used to demonstrate that the summed score have the similar response patterns, thus knowing the total score, one can reproduce all of the responses of each subject (Loevinger, 1948). Loevinger further states that the coefficient is intended to reflect the percentage of all responses which fit the appropriate scale pattern. Stated simply, C of R is the percentage of all responses which are reproducible from the individual's score. In Table 2.1 (page 51), for example, Norstrom & Thorslund (1991) chose to delete the post item, as it produced the most response errors. After the deletion, C of R attains a value of 0.90. The improvement is marginal, but the scale would then fulfil the requirements of a Guttman scale (Norstrom & Thorslund).

A perfect scale relating to functional status could show a cumulative order of age-related decline, but as likely to be expected, a perfect scale is unrealistic due to measurement errors and random variation among individuals. Therefore, C of R is calculated and then evaluated on an acceptable level of measurement error. C of R is formally defined as the ratio of successful reproductions to total responses made on the hypothesis of perfect scalability:

$$C \text{ of } R = 1 - \frac{\text{Total number of errors}}{\text{Total number of responses}}$$

A reproducibility of 90% confirms the existence of a valid cumulative and unidimensional Guttman scale (Asberg & Sonn, 1989). It is also common for researchers to report the Coefficient of Scalability. The C of S serves to prop up C of R reporting in the event that there is a skewness in the distribution of either items or of subjects (Menzel, 1953); if a skewed distribution is present, one can obtain a high coefficient of reproducibility score just by identifying items that were most frequently scored disabled (Asberg & Sonn, 1989). C of S is defined as:

$$C \text{ of } S = 1 - \frac{\text{Total number of errors}}{\text{Maximum number of errors}}$$

A C of S of 0.6 or more is considered the minimum standard of scalability (Lazaridis et al., 1994).

As previously noted, when scales are hierarchically structured using Guttman procedures, the problems experienced when trying to interpret summary scores derived from ordinal ratings are reduced (Doble & Fisher, 1998). However, as early as the 1980s researchers began highlighting serious shortcomings in the Guttman scaling procedure when applied to ADLs. Some of the criticisms included: (1) Guttman is a deterministic model based on the expectation that persons will pass all items that are easier than their ability level and will fail all items that are more difficult. However, this sort of relationship is rarely met, because there will always be some degree of error in reporting on latent traits (Bond & Fox, 2001). Error may relate to, for example, deficiencies in observation methods or individual differences related to item interpretation; subjects may have completely different ideas of how to interpret *some difficulty* with going up steps and *a lot of difficulty* going up steps (Wilson, 2005). Therefore, the relationship to item responses and the construct is better framed as probabilistic rather than deterministic (Kempen et al., 1995). To ensure that a scale conforms to this expectation of clear-cut pass/fail point for each person, the differences between item difficulties must be large (Fisher & Fisher, 1993) Thus, the sensitivity of such scales to small changes in functioning within individuals over time or to small differences between individuals is dramatically reduced (Finch, Kane, & Philip, 1994). Also, the deterministic quality of Guttman reduces the potential to accurately define measurement error. Stated differently, the

item characteristic curves of this model will be a perfect step function, which is problematic because researchers will have difficulty determining whether deviations from the proposed item hierarchy represent error or that the construct is not sufficiently unidimensional (Gillespie, Tenverget, & Kingma, 1987). (2) Rigid hierarchies encountered in Guttman scales have rarely been borne out in either social or behavioural research (Siu, Reuben, & Hays, 1990; Wilson, 1989) “The use of Guttman scales has been found to be severely compromised by the problem of large numbers of response patterns that do not conform to the Guttman requirements” (Wilson, 2005, p.87). (3) Sheehan et al. (2001) reported: Lazaridis and colleagues (1994) studied the scalability of selected ADL items using criteria associated with Guttman scales and found that the Katz ADL fulfilled Guttman’s scaling criteria. However, Lazaridis et al. also found that the Katz hierarchy was one of 360 possible hierarchies, based on permutations of six ADL items. Lazaridis tested all 360 of these hierarchies using the same Guttman scaling criteria and found four additional scoring schemes that performed equally as well as Katz, and found a total of 103 scoring hierarchies that satisfied the minimum standards of scalability according to Guttman. “The fact that there is not a single hierarchical scale, but as many as 103 different hierarchies underlying Katz’s six original ADL items, exposes the disadvantage of a rigid and deterministic hierarchy” (Sheehan et al., 2002, p.843). (4) The coefficient of reproducibility has also been criticised, e.g., Loevinger (1948) asserts that C of R is a highly inefficient statistic, based on only a small fraction of the relevant data. Similarly, Festinger (1947) states, “It is clear that applying a criterion of 90% reproducibility to all attempts at scaling, irrespective of the number of items involved or the number of possible answers to each, leads to false conclusions. In one case where there are many items and many parts to each question, 90% reproducibility might be excellent consistency, in another case 90% reproducibility might represent no better than chance occurrence and be no evidence at all for unidimensionality”.

It appears that the increased interpretive power that is provided by establishing an item hierarchy is suspect under traditional Guttman scaling procedures. As a consequence of the criticisms noted above, in addition to the development of more sophisticated probabilistic methods (i.e., Item Response Theory), Guttman scaling

procedures in health science research have for the most part been abandoned (Vittengl et al., 2006). However, exceptions do exist, with some contemporary research implementing Guttman scaling (LaPlante, 2010). Item response theory (IRT) is potentially more accurate in establishing an item hierarchy, and in some cases, easier to obtain model fit. Also, as a consequence of rather novel methods for addressing measurement error, scales developed or revised through IRT tend to be more precise. Furthermore, some IRT procedures are capable of establishing interval level data. Many item response theory models (e.g., parametric-type) attempt to replicate the precision found in the physical sciences by converting ordinal and qualitative characteristics into interval and quantitative data. A variable at the ‘interval’ level of measurement maintains spacing between numbers which are equal. Thus, an interval variable is one where there is an arithmetical relationship between responses (Glantz, 1992). “The critical difference between the ‘interval’ and ‘ordinal’ level is that all we can say about the latter is that a score of 12 is worse than one of 6” (Tennant, Hillman, Fear, Pickering, & Chamberlain, 1996, p.574). The Guttman model struggles to make such a transformation. The Guttman model assumes that at some point along the disability continuum, a sudden increase (from zero chance to 100% probability of responding to an item) occurs. Teresi, Cross, & Golden (1989) assert that the assumption of the Guttman model is that the underlying trait is *not* continuous, but rather that there are two distinct categories of subjects—those impaired and those not.

2.1.4 One Parameter Rasch models (1PLM)

The most often used item response model is the one-parameter Rasch (Rasch, 1960) model (1PLM) (Farin et al., 2007). The Rasch model is also the most basic IRT model, and is particularly suited for constructing scales (Hopman-Rock, Van Buuren, Kleijn-De Vrankrijker, 2000). Traditional summative Likert approaches to scale analysis focus on the total score of a given instrument, but the Rasch measurement model allows analysis of instruments at the item level (Chiu, Fritz, Light, & Velozo, 2006). Like all IRT models, Rasch establishes, in order to achieve measurement objectivity, the assumption that the probability of a particular response depends on the difficulty of the item *and* the ability of the person (Fortinsky, Garcia, Sheehan,

Madigan, & Tullai-McGuinness, 2000). In the Rasch model (logistic function), the *dependent variable* is the simple probability that person s passes item i , $P(X_{is} = 1)$. The *independent variables* are the person's trait score/latent ability (θ), and the item's difficulty level, β_i . The independent variables combine additively, and the item's difficulty is subtracted from the person's ability, θ . The prediction can be written as follows:

$$P(X_{is}=1 | \theta_s, \beta_i) = \frac{\exp(\theta_s - \beta_i)}{1 + \exp(\theta_s - \beta_i)}$$

The latent trait (θ) is estimated for a model in an effort to obtain a person's responses, while controlling for item parameters. Typically, θ is estimated with maximum likelihood statistics. Thus, to find the appropriate latent trait value, the analyses will first represent the likelihoods of a response pattern under various trait levels and then conduct a search aiming to identify the trait that gives the highest likelihood' (Embretson & Reise, 2002).

Assessing whether items fit the Rasch model is an important criterion. Rasch model residuals in raw form are squared to obtain the outfit and infit mean squares for each person, item and category threshold (Wright & Moc, 2004). The mean square statistics usually consist of different ways of summing squared residuals—where a residual is the difference between a person's actual response and the one predicted by the model. More precisely: "the probability of person n succeeding on item i (P_{ni}) is the expectation of the observed (X_{ni}). The size of the discrepancy (Y_{ni}) between the observed and expected values is evaluated. Thus $Y_{ni} = X_{ni} - P_{ni}$." (Wright & Moc, 2004, p.,11). Chi-square is then used to determine the significance of the discrepancy. Deviations in excess of the expected value can be interpreted as 'noise' or lack of fit between items and the model, whereas values significantly lower than the expected value can be interpreted as item redundancy or overlap (Smith, 2005).

As highlighted previously, an important function of IRT is to allow for the conversion of ordinal data into interval measure. Prior to the conversion of interval measurement, scale items must satisfy criterion of the Rasch model (de Morton, Keating, & Davidson 2008). One such criterion relates to scale unidimensionality, or local independence. Item fit statistics (mentioned above) are an indicator of whether or not each item contributes to the measurement of a single underlying construct. Both item fit and person fit to the Rasch model can be evaluated (Smith, 2005). The unidimensionality assumption asserts that the only factor influencing response behavior is the one common variable θ (latent trait) and random error (Reise & Henson, 2003). Performing poorly on an easy task and to not have increasing difficulty in responding to harder tasks are exceptions, i.e., errors in the perfect scalogram. These exceptions are typically assessed through goodness-of-fit statistics. Items with large negative residual values indicate a high level of predictability in response and therefore information redundancy. Items with large positive residual values indicate an unacceptable level of “noise” in the responses. Within the Rasch model item fit statistics are often the only indicator used to establish scale unidimensionality. It is worth noting, compared with factor analysis, IRT models have considerable theoretical advantages (DeJong & Molenaar, 1987). DeJong & Molenaar further state that whereas factors often are no more than sets of highly correlating items, various IRT methods additionally have a systematic order relationship. Furthermore, the Rasch model, unlike Factor Analysis, is able to model specific response probabilities for each person-item combination (Sijtsma & Molenaar, 2002).

2.1.5 Rasch versus Likert-scoring models

The Rasch-based scoring has been shown to yield greater precision than Likert scoring procedures for purposes of discriminating between age and patient groups (Martin, Kosinski, Bjorner, Ware, MacLean, 2007). Under the Rasch model it is possible to (1) estimate item difficulty as a function of its successful completion, (2) estimate the difficulty of accomplishing each successive step (i.e., item option or category), and (3) estimate a person location corresponding to level of ability (Haley, Fragala, Aseltine, Ni, Skrinar, 2003). The gain in precision is achieved, in part,

because the Rasch-based scoring is a logistic transformation that expands the scale at the upper and lower extremes of the range relative to Likert-based scoring. As a result, Rasch-based scores are more precise than the Likert summative scores when those in comparative groups tend to score at the extremes of the range of physical functioning (Raczek et al., 1998).

Another advantage of the Rasch model relates to the formal confirmation of a hierarchical decline in physical functioning. As mentioned previously, psychometric investigations of ADL scales have primarily supported aspects of reliability and validity (McHorney, 2002). However, evaluating whether functional ability scales form a hierarchical item continuum is frequently overlooked (Haley, McHorney, Ware, 1994). Hierarchy in functional ability can be described in terms of stages along the ability continuum, which summarize similar physical functioning at the same level and distinguish different features across levels or stages (Tao, Haley, Coster, Ni, & Jette, 2008). A person's stage is expected to have prognostic significance and to be useful in the selection of alternative rehabilitation therapies, environmental modifications, or even social behaviours (Cooper, 1980). Cooper proposed a hierarchical classification of normal social behaviour, using Mokken scale analysis to test the following: social dysfunctioning will manifest first in performance of roles (e.g., employment), which involves interaction with people at a greater social distance; with increasing severity, other roles, closer to the patient (e.g., household), will be affected, and in the end, self-care will deteriorate (Cooper, 1980).

With a hierarchical system, all shifts have immediate meaning—each of the levels represents a distinct performance pattern. Summary scores, on the other hand, will not behave this way. In practice, clinicians are seeking less ambiguous guidance on what is happening to the subject (Dejong & Molenarr, 1987). It is the ADL-IADL hierarchy that can best respond to this need. Thus, although additive summary scores may be useful for describing overall function, they maintain less interpretive power. For instance, a hierarchical measure is best positioned to identify a key variable

which may be crucial in terms of understanding or tracking patient status (Morris, Fries, & Morris, 1999)

2.1.6 Discriminatory power

As previously stated the Rasch response probabilities show a nonlinear relationship to total score (the probability of a correct response is a monotonic and increasing function of trait level), and the item characteristic curves, the graphic representation of response probabilities for each item, do not cross. Monotonicity refers to the assumption or requirement that as the latent trait increases so does the probability of a correct response to an item. Both of these requirements are also present in the nonparametric Mokken model, which is discussed below. The non intersection assumption implies that items maintain their difficulty order across score (or ability) categories. Item difficulty corresponds to item location which describes the extent to which items differ in probability across trait levels. When item ICCs do cross, the Rasch model will not be adequate to characterise data. In such cases, items are often deleted, or more complex IRT models are employed, i.e., models that incorporate a discrimination parameter. The Rasch model considers all items to be equally important for measuring, but other models (e.g., two-parameter logistic model) give more weight to better discriminating items (Embretson & Reise, 2002).

2.1.7 Mokken model

The IRT models proposed by Mokken (1971) are called nonparametric because the ICCs are not (unlike Rasch) parametrically defined, and because no assumptions are made concerning the distribution of the latent trait. Mokken's model is used for building hierarchical, unidimensional scales. Like Guttman's model, Mokken analysis assumes an underlying latent attribute is represented by a homogeneous set of items. The main difference between the two is the stochastic or probabilistic nature of the Mokken model compared to the deterministic nature of the Guttman model (Dejong & Molenaar, 1987). The Mokken model of monotone homogeneity is based on the assumptions of unidimensionality, local stochastic independence, and monotonicity in the latent attribute (Watson, Deary, & Austin, 2007). Unidimensional measurement, as discussed previously, states that within a specified population,

subject responses can be explained by a single underlying attribute. This attribute is measured on a scale denoted by θ . Local stochastic independence implies that the response behaviour of a person on an arbitrarily selected item g is not influenced by his or her responses on previous items, nor will it affect response behaviour on subsequent items. For instance the two items, one's ability to walk several blocks and the ability to walk one block, are likely to violate the assumption of local independence. The assumption is a logical consequence of unidimensionality, but the reverse is not true (Meijer, Sijtsma, & Smid 1990). The assumption of monotonicity in the latent ability specifies that a higher ability value implies an increasing probability of responding positively to an item (Meijer et al., 1990).

The Mokken scaling procedure often begins with inspecting the scalability of individual items (H_i), which are used to summarize item discriminatory power. The H_i coefficient must be nonnegative and its maximum value is equal to one. H_i is used to assess discrimination power; an item with values between 0 and 0.3 are thought to have weak discrimination power. A high H_i value suggests that a task fits well with the rest of the scale items. It also differentiates between people with low probabilities and those with high probabilities of impaired functional ability as measured by the whole scale (Embretson & Reise, 2000). Less discriminating items have response probabilities that are less responsive to changes in trait level (Embretson & Reise, 2000.) The H_i value is considered a convenient summary of item functioning. However, removing ineffective items solely on the basis of H_i values is not recommended. A diagnosis of the misfit of item i should also include the investigation of the Item Response Functions (IRF) of this item. ICCs or IRFs can be used to identify possible causes of a low H_i value: (a) an almost flat IRF; (b) an irregular IRF that jumps up and down across θ ; (c) a single-peaked IRF (Sijtsma & Molenarr, 2002).

2.1.8 Mokken scaling versus the Rasch model

The Rasch model and the Mokken model are both unidimensional cumulative models: both models assume that there is only one latent trait underlying the answers and that the probability of a positive or a correct answer for each item is a non-

decreasing function of this latent trait value (i.e., the ICC is non-decreasing). The two models differ mainly in the assumptions they make about the shape of the response probabilities of person and items (Junker, 2001). Also, the two models differ in the way they estimate response probabilities (ICCs or IRFs): the Mokken model, unlike Rasch, does not make use of numerical estimates of θ . Instead, the Mokken model replaces θ with the total sum score of all items except item i being investigated ($k - 1$ items). This total score is referred to as the rest score ($R_{(i)}$). Thus, IRFs are estimated from the data by defining persons with $R_{(i)}$ and then calculating the fraction of persons with a particular $R_{(i)}$ who have a score of 1 on item i (1 denoting successful completion of a task within a dichotomous scoring system) (Sijtsma & Molenaar, 2002). Perhaps most important is the notion that Mokken scales provide a more flexible, and thus are likely to provide a framework in which the data fits the model, as compared to Rasch model.

2.1.9 The Two-parameter parametric model (2PLM)

The one-parameter Rasch model (1PLM) appears to be the most frequently used IRT method when interpreting ADL data. Yet, both the Mokken model and the two-parameter parametric (2PLM) model offer increased interpretive power. Like the Mokken scaling procedure, the 2PLM model is able to calculate both item difficulty and item discriminatory power. Unlike the Rasch model, the 2PLM and Mokken models define unique discrimination parameters for each item rather than a single slope estimate for all scale items. In the 2PLM model, item discriminations are related to the biserial correlations between item responses and total scores. When item discrimination is included in an IRT model, trait level estimates depend on the specific pattern of success and failures in the item set. Thus, in the 2PLM model items do not have equal weight in estimating trait level (Embretson & Reise, 2002). Furthermore, if the 2PLM model produces a broad range of slope values (i.e., discriminatory power estimates), it is likely that the Rasch 1PLM model will not fit the data well (Cabrero-Garcia & Lopez-Pina, 2008).

The 2PLM model is generally thought to be more precise than the Mokken method. Embretson and Reise (2002) indicate that the 2PLM models should be favoured if parameter estimates must be very accurate, and yet fit statistics are barely reliable for scales containing few items and small samples (Cabrero-Garcia & Lopez-Pina, 2008). Consequently, Mokken procedures should be implemented if the sample size is relatively small, or to offer a more flexible framework when 2PLM parametric models do not fit well with particular populations and item characteristics (Embretson & Reise, 2002). Sample size is an essential criterion when applying IRT models. A sufficient spread over the range of outcomes and sufficient sample size will allow for the estimation of all model parameters; for 2 PLM models, sample sizes of 500-1000 are probably sufficient (Tsutakawa & Johnson, 1990). Similarly, Reise and Yu (1990) recommend approximately 500 subjects.

2.2 Summary

In conclusion, the most common procedure today is the summing (Likert-scoring) of item ratings to derive a total ADL score (Ward & Murray-Ward, 1998). No doubt the allure of this aggregated method relates to its simplicity. Also, ADL-IADL scales often perform well in factor and internal consistency analyses. Additionally, summative scores have been shown to yield similar levels of validity when compared to Rasch scoring (dichotomous & trichotomous options) (Vittengl et al., 2004). However, even with the most basic IRT model (Rasch 1PLM), summative scores appear to be less valid. For example, Vittengl et al., found that increasing the number of item options beyond three resulted in a decrease in sensitivity for summed scores (as compared to Rasch) at the high and low end of the ability continuum. This suggests that the association between Rasch and summed scores was monotonically increasing, but summed scores increased less per Rasch interval in the lowest and highest ranges of the scale compared with the middle (Fortinsky et al., 2003).

Findings of similar validity for some types of Rasch models do not bear on other strengths of IRT as applied to ADL-IADL instruments. Namely, (1) the ability to form item hierarchies; (2) to streamline testing by eliminating items that are much

too easy or difficult; (3) examining the characteristics of individual items, and determining if polytomous scoring categories work as intended; (4) to assess discrimination power of individual items; (5) the ability to test invariance of items across external groups (i.e., differential item functioning). “Thus, a vigorous protocol is used to test what, in effect, is the internal construct validity of the scale” (Davidson, 2008, p.223).

In the next chapter I set out to investigate whether IRT principles could be used to improve our understanding of how functional status presents itself in relatively healthy older adults. That is, can we revise traditional ADL measures (initially developed for very low functioning populations) to be more sensitive so that they can be used to detect early signs of disability in relatively high functioning community-dwelling older adults, i.e., detect preclinical disability. My investigation took the form of a systematic review. The review was relatively narrow in that I only included manuscripts that employed IRT to assess instrument and person properties. Furthermore, the samples used in the assessments were limited to community-dwelling subjects. I pay particular attention to the identification of construct under-representation. Construct under-representation refers to gaps along the disability continuum that make accurate assessment more difficult. Chapter three also addresses the potential improvement in validity as it relates to establishing a formal hierarchy of functional decline. This topic is particularly relevant to this thesis, in that, such hierarchies allow for the use of self-reported functional status measures to be used as predictors of disability, as well as the more traditional role of outcome measure.

Table 2.1 Guttman analyses of seven ADL-IADL items (n=421)

Percentage in each scale-score dependent in:								
Scale-score	Bed-making	Cooking	Post	Bathing	Food shopping	Laundry	Cleaning	%
7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	7.0
6	57.3	85.3	72.0	96.0	89.3	100.0	100.0	4.9
5	22.7	62.7	57.3	64.0	97.3	96.0	100.0	4.9
4	0.0	38.8	44.6	67.8	90.1	60.3	98.3	8.0
3	0.0	38.2	47.3	22.9	51.1	55.7	84.7	8.6
2	5.2	0.6	22.1	31.8	15.6	40.9	83.8	10.1
1	0.0	0.4	3.4	7.6	9.9	17.1	61.6	17.3
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	39.1
%	11.4	20.8	23.8	26.8	31.1	33.4	51.1	

Note: table reproduced from Norstrom & Thorslund (1991).

3

Evaluating the sensitivity and accuracy of instruments used to assess functional status in community-dwelling older adults: a systematic review

3.1 Background

In the U.S., the number of those aged 65+ in the year 2000 was approximately 35 million. In 2050, this figure is expected to rise to nearly 82 million (Wiener & Tilly, 2002). Such forecasts have prompted gerontologists and geriatricians to consider more seriously prevention-type models, with an emphasis on the earliest stages of functional decline. Increased interest in the maintenance of function and prevention of disability has led to relatively new diagnostic criteria, such as symptoms of frailty or preclinical disability. The utility of identifying individuals who are ‘high risk’ for future functional decline rests on the notion that it is potentially an easier state to reverse than overt disability (Richardson, et al. 2008). Intervention programs designed to prevent functional decline in older adults show that participants with relatively good functional status or moderate frailty are those who benefit the most from these programs (Gill, et al. 2002). However, ‘prehabilitation’ strategies necessitate the use of assessment measures that exhibit a high degree of sensitivity. Standardised tests of physical performance have been employed with increasing frequency in recent years, presumably to meet this demand for greater sensitivity (Coman, & Richardson, 2006).

Activities of Daily Living (ADL) (Katz, Ford, Moskowitz, Jackson, & Jaffe, 1963), and Instrumental Activities of Daily Living (IADL) (Lawton, & Brody, 1969), were developed to assess capabilities relating to the maintenance of self and lifestyle, which often includes self-care, keeping one’s life-space in order, and obtaining resources (Sonn, 1996). At one point Activities of Daily Living and Instrumental

Activities of Daily Living were the most widely used measures in gerontology (Guralnick, & Simonsick, 1993). The popularity of these self-report instruments was due to their association with a multitude of variables, such as morbidity, mortality, risk of institutionalization, and need for care (Wilms, Riedel-Heller, & Angermeyer, 2007). However, compared with performance-based measures (e.g., walk time), ADLs and IADLs generally display weak face validity, reproducibility, and sensitivity to change (Guralnick, Branch, Cummings, & Curb, 1989). Also, as the emphasis has changed toward early detection in community-dwelling older adults, for whom dependency in self-reported ADL-IADLs is uncommon, researchers often have to cope with large ceiling effects, in which greater than 90% of subjects endorse no 'difficulty' or 'dependency' on ADL tasks (Reuben, et al. 2004). It has been proposed that the relative standing of ADL-IADLs could be enhanced by improving construct validities to levels that are at least equivalent to those of physical performance measures (Kinugasa & Nagasaki, 1998). Enhancements of this nature have progressed relatively slowly. The justification for improving construct validity in ADL-IADLs, rather than abandoning them in favour of performance measures, can be found in two observations. First, there is evidence that self-reported ADL-IADLs and performance based measures are comparable, but usually measure different aspects of functioning (Katz et al. 1963). Second, combining information from self-report and performance measures has been shown to increase prognostic value, particularly in high-functioning older adults (Reuben, et al. 2004).

In addition to evaluating content validity associated with ceiling effects for community-dwelling populations, this review will investigate construct validity as it relates to construct under-representation. In particular, item response theory (IRT) models are capable of transforming non-linear ordinal scores into linear measures. "With the priority placed on establishing interval units of measure, the investigator derives complementary tools for understanding the nature of a scale's meaning and, more importantly, provides a substantive context within which an individual's score on a scale may be interpreted" (Haley, et al. 2004, p.52). Establishing interval level units allows one to identify important features of the construct that have been excluded, which results in measurement gaps. This means that there may be uneven

rates of change in the construct being measured. For instance, an increase in a 10-point scale can represent different amounts of improvement at different parts of the functional status scale; it might be more difficult for a person to improve from 9 to 10 than from 4 to 5 (Liao & Campbell, 2004).

Construct validity for functional status scales can also be enhanced by formally confirming a hierarchy of decline; for example, by supporting or refuting the expectation that ‘Stepping over obstacles’ is a more challenging task than ‘Walking over a level surface’ (Chiu, Fritz, Light, & Velozo, 2006). Establishing a hierarchy of functional decline tells more than the typical simple summation of functional loss, and may have predictive value to the clinician monitoring older adults: if the sequence is accelerated or out of order it may indicate the need for interventions (Daltroy, Logigian, Iversen, & Liang, 1992). Furthermore, hierarchical scales can be especially useful to summarize the level of disability in epidemiological studies, and can act as a discriminative index for measuring change in a longitudinal research design (Kempen, Miedema, Ormel, & Molenaar, 1996). IRT-based transformations allow for items to be ranked unequivocally on a hierarchy based on item difficulty, ranking items from easiest to most difficult (Martin, Kosinski, Bjorner, Ware, & MacLean, 2007). Ordering items or tasks by group mean scores does not imply that this ordering also holds at the individual level. “Any set of items can be ordered by item mean scores, but whether such ordering also holds for individuals has to be ascertained by means of empirical research. Only when the set of items has an invariant item ordering (IIO) can their cumulative structure be assumed to be valid at the lower aggregation level for individuals” (Ligtvoet, van der Ark, te Marvelde, & Sijtsma, 2010, p.579).

We are concerned with examining performance in relatively high functioning community-dwelling older adults, which requires novel approaches to scale construction (e.g., types of items), along with advanced methods for scrutinising such scales. Most practitioners still limit their scale analysis to methods relating to classical test theory (CTT). For instance, employing the classical standard error of measurement based on Cronbach’s Alpha and the standard deviation of raw scores. A

more recently developed and very powerful alternative to classical test theory is item response theory (Hulin, Drasgow, & Parsons, 1983). Item response theory (IRT) is capable of providing more detailed information about the standard error of measurement. One fundamental advance represented by IRT, compared with CTT, is that IRT models explicitly estimate the joint relationship between person properties (ability along the latent trait) and item properties with the same model (Reise & Henson, 2003). That is, log-odd units are computed for both items and person separately and placed on a common scale (Schumacker, 2004).

Item response theory (IRT) derives from what is called modern test theory, and is one of the methodologies that have resulted in what Embretson (1996) calls “the new rules of measurement.” IRT has several main advantages over classical test theory: 1) IRT is less concerned with scores on sets of items (test level scores) and more concerned with the responses to individual items. Although tests have always been composed of multiple items, IRT takes a much more item-level focus than CTT, which tends to focus more on test-level indices of performance (e.g., the overall reliability coefficient, or standard error, of a scale); 2) IRT enhances interpretive power by providing measurement precision that varies with a person’s ability level (Hambleton, Swaminathan, & Rogers, 1991). The discrepancies between observed and true scores indicate how much the test score differs from the true scores, and are summarized by the standard error of measurement; better reliability estimates result from high precision or relatively small measurement error (Alagumalai, Curtis, & Hungi, 2005). This information (i.e., error that varies by person performance) can be used to identify the most sensitive part of the instrument or scale under investigation (Wilson, 2005). Whereas in CTT a single number (e.g., the internal-consistency reliability coefficient, or the SEM based on that reliability) would be used to quantify the measurement-precision of a test, a continuous function is required in IRT to convey comparable data (Harvey, & Hammer, 1999). 3) Test scoring: one reason given as to why the psychometric properties of self-reported ADL-IADLs can be insufficient pertains to the ordinal nature of Likert scoring methods. Likert summative scoring is the most common scoring method for ADLs (McHorney, Haley, & Ware, 1997). This traditional aggregate method of

scoring computes a raw total score by summing responses to individual items. The sum score of the responses to the items is the estimator of the position of the patient on the continuum. The rationale behind this procedure derives from classical test theory (Van Alphen, Halfens, Hasman, & Imbos, 1994).

Despite the popularity of the aggregate scoring method, there are well-established problems with raw scale scores that make them difficult to interpret (Reise & Henson, 2003). One problem pertains to weighing each item equally; the total score method assumes that each item or symptom on the scale represents an equal level of severity, which is almost never true (Gibbons, Clark, Cavanaugh, & Davis, 1985). Revised ADL-IADLs, through the use of Item Response Theory (IRT), avoid the pitfalls of aggregated approaches to self-reported disability. In contrast to traditional summative scoring methods, IRT models meet the conceptual requirements of order and additivity (McHorney, Haley, & Ware, 1997). The two methods (i.e., IRT vs. Likert scoring), with respect to difficulty ranks, can diverge considerably. For example, it has been demonstrated, within a 16-item scale, five Likert items scores differed by three or more ranks compared to Partial Credit (Rasch model) scores (Fortinsky, Garcia, Sheehan, Madigan, & Tullai-McGuinness, 2003).

The goal of this systematic review is to identify manuscripts that use Item Response Theory to revise or develop ADL-IADL scales used for community-dwelling older adults. These revised scales should: (i) assess internal validity (cause and effect) by formally confirming a hierarchy of functional decline; (ii) enhance content validity, by reducing ceiling effects to the threshold of 15%, or even 10%, i.e., only 15% of the sample will exhibit no traces of the latent construct—namely, disability; and (iii) quantify construct under-representation (i.e., gaps in coverage) by converting the raw aggregated disability score into interval level measurement. The by-product of the aforementioned goals will be the identification of ADL-IADL instruments that are highly sensitive to early declines in functional status, and more accurate in detecting change over time. Lastly, this review is not concerned with establishing the superiority of one method over another (i.e., item response theory vs. classical test theory) in relation to scale analysis.

3.2 Methods

3.2.1 Data sources

Published studies were identified through searches of MEDLINE (from its inception in January 1966 until November 2008), PSYCINFO (1872 until November 2008), EMBASE (1980 until November 2008) and CINAHL (1981 until November 2008) databases. Keyword, title and abstract information were used. The main search terms included 'functional decline' or 'function*' (the symbol is used for identifying all words starting with function, e.g., functional, functions) status' or IADL or 'instrumental activities of daily living' or ADL or 'activities of daily living' or BADL or 'basic activities of daily living' or 'personal activities of daily living' or 'functional disability' or 'functional tasks' or 'loss of independence' or disabled or disabilit* or 'functional impairment' AND 'cumulative structure' or 'scale construction' or 'guttman scaling' or mokken or rasch or uni-dimensional* or hierarch* or unidimensional* or IRT or 'item response theory' or 'patterns of functional decline' or scalogram or 'cumulative order' or 'one dimensional' or 'psychometric properties'.

Figure 3.1, on page 74, depicts the flow chart for this review. After selecting 106 articles for full review, the reviewer examined the reference sections of these articles, which resulted in a total of 12 articles that required a full review. The initial search criteria included 'all languages'. Unpublished studies, dissertations, theses, book chapters or manuals, and studies published in non-peer-reviewed journals were not considered for the review.

3.2.2 Inclusion and exclusion criteria

Generally, reports were included in this review if they described instruments with face validity for measuring disability, and thus closely reflect the fourth dimension of the Nagi (1965) model (difficulty doing activities of daily life, such as employment, household management, leisure activities, personal care, etc). The scales in this review will most likely resemble traditional Instrumental Activities of Daily Living

(Wilms et al. 2007), but will also, to a lesser degree, incorporate Basic or Personal Care Activities of Daily Living, as well as functional tasks (e.g., bending and kneeling, or walking outdoors). The latter more closely resembles the third dimension of the Nagi model. Scales were required to be generic measures; that is, should not be disease specific. The authors of this review chose to limit subject inclusion to those individuals 50 + years, with a sample mean age of 60 and above. Papers needed to scrutinize ADL-IADL performance with item response theory methods or Guttman scaling procedures. Reports that were primarily concerned with how broad domains of functioning, such as mobility, instrumental activities, and self-care activities form a hierarchy, while neglecting to assess item level functioning were not included; these studies presume a multidimensional structure to disability, and thus assess a hierarchy between domains. Manuscripts examining functional decline with a Medicare sample *were* included in this review, but were interpreted with caution, as these sample populations were generally more severely impaired than other community-dwelling samples. Studies using proxy reports were not included due to previous findings indicating a discrepancy between self-report and proxy ADL-IADL measures (Rogers, & Miller, 1997). Despite the inclusion of manuscripts that utilised Guttman scaling procedures in our initial search criteria, in the end these manuscripts were excluded from the review. This was done for one of two reasons: 1) there is a large body of evidence asserting the inferiority of Guttman methods compared with more advanced IRT procedures (see Supplementary material); and 2) many first generation functional status measures (i.e., Basic-ADLs) employed Guttman scaling procedures. Scales strictly examining Basic-ADLs are less relevant to this review because they are ineffective in assessing community-dwelling older adults, e.g., as few as 8% of community-dwelling subjects reported ADL dysfunction (Branch, Katz, Knipmann, & Papsidero, 1984).

If the inclusion exclusion criteria could not be ascertained from examining the title or abstract of the manuscript, then the article was selected for full review. In the full review of articles, considerable attention was paid to the methods and results sections. Again, of primary interest were the sample of subjects (i.e., were they community-dwelling subjects) and whether IRT procedures were implemented. My

rationale for selecting scales that just used IRT methods relates to the notion that scales revisited through IRT procedures can provide more detailed information (as compared to scales solely examined by means of classical test theory) about an older adult's level or degree of disability. With regard to extracting information, it was particularly important to extract information pertaining to scale reliability, i.e., could the observed scores be reproduced if a parallel scale was readministered to the subject. The manuscripts should quantify item difficulty through parametric or nonparametric means. Lastly, the manuscripts will provide some evidence that unidimensional construct has been observed.

3.2.3 Reliability

Scale reliability was measured in one of four ways: Item or Person Separation Index, Item or Person Separation Reliability, Test Information Function, and Rho Coefficient. The Test Information Function (TIF) represents the inverse of *standard error of estimation*. This standard error of estimation serves the same role as the standard error of measurement in classical test theory, except that the former statistic can vary for each examinee (Hambleton, et al. 1991). The TIF can be used to identify the most sensitive part of the instrument or scale (Wilson, 2005), Item reliability and separation statistics refer to the ability of the test to define a hierarchy of items along the measured variable, and the higher the number the more confidence we can place in the reliability of item placement across other samples or test administrations (Bond & Fox, 2001). A similar principle applies to the *person* reliability and person separation index, i.e., replicability of person ordering and sufficient spread of person ability across the continuum. The reliability of item difficulty or person ability is interpreted on a 0 to 1 scale (similar to the way in which Cronbach's alpha is interpreted). These reliability estimates can be transformed to an item or person separation index, which reflects the number of standard errors of spread among the items or persons; higher separation indicates a scale that covers a wider range of the construct being measured (Bond & Fox, 2001). In assessing the separation index, the value should be at least 2 to obtain the desired reliability coefficient of .80. A *person* separation index of 2.0 indicates that the sample can be separated into at least three distinct groups (Fisher, 1992) and an *item* separation

index of 2.0 means that the items can be divided into three distinct levels of ability (Arnadottir, & Fisher, 2008). For the nonparametric Mokken scaling, Rho is used to define scale reliability, and is an internal consistency coefficient comparable to Cronbach's alpha (Moorer & Suurmeijer, 1994). Most theorists agree that a Rho over .80 is desirable, and a Rho over .70 is a minimum requirement (Kempen & Suurmeijer, 1990).

3.2.4 Validity

Of the four types of validity outlined by Cronbach and Meehl (1955), this review will be most concerned with examining construct validities for each paper selected, as well as one aspect of content validity—namely, ceiling effects. An important aspect of construct validity is the trustworthiness of score meaning and interpretation (www.rasch.org/rmt/rmt221a.htm.) It has been proposed that two major threats to score meaning and interpretation are *construct-irrelevant variance* and *construct under-representation* (Messick, 1989). The former reflects unrelated sub-dimensions that are irrelevant to the construct being measured (e.g., disability), and the latter refers to the exclusion of important features of the construct (i.e., gaps in continuum coverage). Construct under-representation can be observed for parametric IRT models that provide interval level data. Because health status instruments are summed scores and typically include zero it has been common to treat them as continuous variables with ratio or interval characteristics. However, definition of a zero point is arbitrary and instrument dependent (Messick, 1989). Furthermore, if the distance between items is not equally spaced, a segment change in an area of the scale with high item density will produce a greater numerical gain than a segment change in an area of the scale with low item density, despite the change being of equal magnitude. Typically, equally spaced interval units are derived by converting the raw score percentage into a success-to-failure ratio. Then the natural log of this odds ratio is computed.

Establishing a formal hierarchy of decline, or invariant item ordering (IIO), should enhance construct validity. In Likert scale models no strict item hierarchy is hypothesised or defined and priority is given to internal consistency (Messick, 1989).

With IIO, the order of the items in terms of difficulty should be the same for all respondents whatever their latent trait value (Sijtsma, & Junker, 1996). Ligtvoet et al. (2010) convey that IIO is strong requirement in psychometrics, and that researchers wrongly assume that fitting any IRT model implies that the items have the same ordering by difficulty for all subjects. Furthermore, previous research has shown (Sijtsma, & Hemker, 1998), rather surprisingly, that only restrictive polytomous IRT models provide IIO, i.e., rating scale models (Andrich, 1978; Muraki, 1990). With regard to dichotomous-item tests, Sijtsma and Junker (1996), demonstrated that the Rasch model (Rasch, 1960), and the double monotonicity Mokken model (Mokken, & Lewis, 1982), can also be used to establish IIO. The Mokken model for polytomous items also provides diagnostics for establishing IIO; in this case the criteria for IIO are met when the percentage of negative coefficients at the level of the individual subjects (H_a^T) is less than 10% and the coefficient for total set of subjects (H^T) is at least .30 (Roorda, Roebroeck, van Tilburg, Lamkhorst, & Bouter, 2004).

Content validity assesses whether the items measure what they claim to measure, and also if they measure the full range of the construct, which is discussed in terms of floor and ceiling effects (Salomonsson, Ahlstrom, Dalen, & Lillkrona, 2009). These effects are the results of an item(s) clustering in the highest or lowest result group. The distribution of the results in the different review scales are presented and evaluated. The floor and ceiling effect is also considered important for the analysis of responsiveness. Floor and ceiling effects are presented in terms of responsiveness because they indicate limits to the range of detectable change, beyond which no further improvement or deterioration can be observed (Salomonsson et al. 2009). A maximum of 15% for any given sample has been proposed as the reasonable limit of ceiling or floor effects, with some investigators suggesting a ceiling threshold as low as 10% (de Morton, Keating, & Davidson, 2008). Bruce & Fries (2003) raise an important point concerning ceiling effects that is worth mentioning here. These authors suggest that, for example, if 10% of a sample of patients with rheumatoid arthritis have scores of zero on an ADL scale, this may not be an indication of a ceiling effect. But rather, it can be interpreted to mean that if 10% of patients report

no difficulty for any of the ADL items, then 10% of these adults with arthritis have no disability. Within the context of the Bruce & Fries commentary, this chapter is not concerned with identifying the construct of disability. The sample of subjects examined throughout this thesis is relatively healthy community-dwelling older adults, thus we could expect that only 50% of the sample might endorse ADL disability. However, because we limit our investigation here to the detection of disability *symptoms* in community-dwelling older adults, the construct under investigation would more accurately be described as preclinical disability. Because some individuals avoid disability, we would not expect a ceiling effect of zero%; it has previously been observed that as much as 20% of persons aged 95 and over have been shown to require no assistance with ADLs (Spector et al., 2000). Given this demographic information, it would appear unreasonable to expect ceiling effects to fall below 15%.

3.3 Results

3.3.1 Articles close to inclusion

Of the 106 articles selected for full review, six articles were excluded with some hesitation. Below is a list of articles that were very close to being included in the final list of ‘review articles’, but were ultimately excluded. All authors responded, but indicated that additional information was unavailable. 1) Avlund, Shult-Larsen, & Kreiner (1993) was excluded due to data unavailability, specifically logit calculations and reliability coefficients. Avlund, Kreiner, & Shultz-Larsen (1993) and Avlund, Kreiner, & Schultz-Larsen (1996) were also excluded because logit information was unreported. McHorney (2002), required reliability and item fit statistics for the community-dwelling sub-sample. In Finalyson, Mallinson, & Barbosa (2005), the reliability coefficients, logit estimates, and fit statistics for community-dwelling subjects are not clearly separated from nursing home subjects or those receiving in-home services. Finally, for Cabrero-Garcia & Lopez-Pina (2005), the analysis was solely conducted between gender groups. However, despite the insufficient information provided, several of these manuscripts will receive further attention in the discussion section of this manuscript.

Details of the twelve studies that met the full inclusion/exclusion criteria are listed below in Table 3.1 on page 75. The table includes a number of factors thought to influence scalability, such as sample characteristics (Kempen, Myers, & Powell, 1995). We chose to highlight, in bold type, the samples that were disproportionately female or male because gender has been shown to significantly affect item ordering (Fleishman & Spector, 2002).

3.3.2 Reliability

A primary advantage of IRT is the extension of reliability. Traditionally, reliability (i.e., the degree to which a scale is free of measurement error) has been used to assess a scale's *average* reliability. IRT on the other hand, with the use of the information statistic researchers can determine how precise a scale is at various ranges of the latent trait (Embretson, 1996). Dubuc et al. (2004), was the only manuscript to report a test information function, with a maximum score of approximately 4.5, which yields a standard error of .47. Despite the information curve being relatively flat and evenly distributed across the disability continuum, 4.5 is a rather modest value for this indicator of precision (Baker, 2001). Hambleton et al. (1991) suggest that a TIF ≥ 10 is preferable. At any point on the latent variable, the standard error of a person estimate (on the complete set of items) is the inverse square-root of the TIF, so that a TIF of 10, person measure standard error = 0.32. Table 1 reports four different methods for assessing scale IRT-type reliability: Item or Person Separation Index, Item or Person Separation Reliability, Test Information Function, and Rho Coefficient. Several studies reported person reliability estimates, without reporting item reliability, i.e., Sheehan et al. (2002), and Spector & Fleishman (1998), both reported a person reliability estimate of .88. These values indicate that the scale can differentiate persons on the measured variable (i.e., disability), and that one can place confidence in the reproducibility of placements. However, these values provide only half the picture, particularly if we are concerned with confirming a hierarchy of functional status items. Haley et al. (2002) and Jette et al. (2002), administered the Late-life FDI and recorded an item separation index of 10.1 and 9.39 respectively, which is well beyond the minimum requirement, and thus

we can be confident that the scale provides an adequate number of statistically distinct difficulty strata with which to measure persons. Of the four manuscripts that employed Mokken scaling, all except Watson et al. (2010) were far above the minimum requirement of 0.70. The Watson et al. functional status scale exceeded the minimum requirement for Rho, but fell short of the desired .80 mark.

One manuscript, Schumacker (2004), that met the inclusion/exclusion criterion for this review was ultimately rejected (and not included in Table 3.1) because the reliability of this instrument was thought to be poor, so that score interpretation or inferences were impeded. The low person reliability value indicates that older adults are not responding in a consistent fashion across the set of 9 activity items for this scale. There appears to be an adequate person separation index, which means that there exists a large enough spread of ability across the sample so that the measures adequately reflect functional ability. However, the low person reliability suggests that the person ability estimates are not well targeted by the item pool. In most applications of IRT, reliability is estimated for both persons and for items. The Schumacker manuscript supports the utility of reporting both person and item statistics.

3.3.3 Construct validity

Seven scales from this review were able to establish interval level measurement using parametric IRT procedures. This enabled greater accuracy when considering change scores as well as identifying construct under-representation. All scales presented with relatively large gaps in coverage, with the exception of McHorney and Coen (2000). Table 3.2 (pages 76-77) provides a summary of all the scales from this review that report interval level data. A relatively common method used to evaluate the distance between item calibrations is to perform a *t* test between successive pairs of items along the logit scale (Bond & Fox, 2001). A gap in the item difficulty measure, which is defined as a significant *t* test for the difference between the measures of two successive items, is evidence of discontinuity in items (Liao & Campbell, 2004). However, when commenting on distances, one often needs to consider each authors definition of “difference” combined with their

sample size and the structure of specific rating scales. And yet, some guidelines or standards have been proposed: a minimum spacing of .15 logit units should ensure that items are distinct from each other, (Wolfe & Kong, 1999), and a ‘gap’ beyond .30 logits might signal the need for additional items to avoid construct under-representation (Jackson, Draugalis, Slack, Zachry, & D’Agostino, 2002). We limit our commentary of gaps to the percentage of interval space that exists between adjacent items.

The Spector and Fleishman scale (1998), covers a logit range from -.83 to 1.61. There is a large gap in coverage between ‘Shopping’ and ‘Doing laundry’, which makes up 26% of the scale coverage. There is another gap (21% of the scale range) between ‘Telephoning’ and ‘Incontinence help’. In Haley et al. (2004), the coverage is relatively even, except for a large gap between the most difficult item, ‘Run half mile’, and the second most difficult item, ‘Hike several miles’; the gap covers 22% of the scale. In Sheehan et al. (2002), there exists one large gap between the two least difficult items, i.e. ‘Lift a full cup or glass’ and ‘Turn faucets on and off’. The gap in coverage represents 21% of the scale range. There is another gap between the two most difficult items, which reflects 15% of to total scale coverage. In Fortinsky et al. (2003), we find a 13% gap between ‘Grooming’ and ‘Ambulation’, a 13% gap between ‘Transferring’ and ‘Feeding’, as well as a 10% gap between ‘Transport’ and ‘Bathing’. Dubec et al. (2004), records two large gaps at the top and bottom of the scale which occurs between ‘Vigorous activities’ and ‘Walk one mile or more’ (21% of the scale range), as well as between ‘Walk one block’ and ‘Bath or dress self’ (29%). Jette et al. (2009), also records two large gaps in coverage, one between ‘Active recreation’ and ‘Volunteer job’ (range of 24%), as well as a gap between ‘Personal care needs’ and ‘Take care of health’ (22%). McHorney and Cohen (1993), use the more complex 2-parameter scaling method, along with equating methods which allows for a large number of items (i.e., 166) to be placed on an interval scale. It is important to note that the Mokken scaling employs nonparametric procedures which do not produce a numerical estimate of item difficulty, but rather ranks items by the proportion of correct responses to an item.

3.3.4 Confirming a hierarchy

It should be noted that the number of scales that accurately report invariant item ordering is somewhat limited. This is because only two parametric models from this review are thought to imply invariant item ordering, the dichotomous Rasch model and the polytomous rating scale model (Sijtsma & Junker, 1996; Sijtsma & Hemker, 1998). The nonparametric Mokken model, when reporting the H^T coefficient, is also capable of confirming IIO (Sijtsma & Molenaar, 2002). Table 3.3 on page 78 depicts scales that report invariant item ordering, thus formally confirming a hierarchy of functional decline. As expected, the Basic or Personal Care ADLs represented the least difficult items, or stated differently, difficulty with these items reflects the highest degree of subject severity. Interestingly, tasks that measure dexterity or fine motor skills (e.g., tie a knot or hold a glass) appear to reflect a greater level of severity than some personal care ADLs, such as bathing and dressing. Due to the limited number of scales from this review that are capable of establishing IIO, common items between scales were relatively few. However, if the ‘Up and down stairs’ item from Watson et al. is most similar to the ‘3 flights of stairs inside’ item from Haley et al. then we observe a common 3-item hierarchy for these two sales (i.e., stairs item followed by ‘Get on a bus’, followed by ‘Reach overhead’).

3.3.5 Content Validity

Four of the twelve scales were exceptional in reducing ceiling effects: Kempen and Suurmeijer (1990) reported 5% of subjects at the ceiling level; Fortinsky et al. (2003), also reported a ceiling effect of 5%; Haley et al. (2002) and Jette et al. (2002), observed a ~1% and 0% ceiling effect, respectively. However, it would appear that the success of Kempen and Suurmeijer (1990), and Fortinsky et al. (2003), has more to do with sample characteristics than item or task difficulty. Both scales were categorised in Table 3.1 as having the ‘least healthy’ samples of older adults. This line of reasoning is confirmed by the fact that the bathing personal care ADL appears in the top 3rd of most difficult items for the Fortinsky et al. (2003) scale. Similarly, in the Kempen and Suurmeijer (1990), scale ‘climbing a flight of stairs inside’ appears in the top 3rd of most difficult items, but this is a relatively easy

mobility items when compared to the mobility hierarchy presented in Haley et al. (2002).

With the exception of Schumacker (2004) who found that 70% of their older adults reported an inability to perform 7 out of 9 activities due to fear, most of the floor effects were negligible. Thus, our results are primarily concerned with the identification of ceiling effects. Kempen et al. (1995), found that 85% of the sample could manage the most difficulty item, 'Going up & down stairs'. Spector and Fleishman (2008) began their study by restricting their sample to those subjects that were functionally disabled in at least one task (4463 to 2977). Thus the ceiling could be considered to include 32% of subjects, which was very similar to that reported in Watson et al. (33%). Kempen et al. (1996), reported ceiling effects for 44.8% of the sample (n=2144) and 8.4% of the sample (n=403) scored ≥ 36 on the GARS (theoretical range of 18-72). Sheehan et al. (2002), also reported a very large ceiling effect, n= 2079 (46.9%). Dubuc et al. (2004) indicated a ceiling effect of 16%. McHorney and Cohen (1999) reported that ~ 15% of their subjects had no difficulty with the six largest location parameter estimates, i.e., the 6 most difficult items. Fortinsky et al. (2003) and Kempen and Suurmeijer (1990), reported similar ceiling effects; in Fortinsky 5% of subjects reported no disability and Kempen & Suurmeijer (1990) found that 5% of subjects reported no problems with the most difficult item. Jette et al. (2003) and Haley et al. (2002), recoded the lowest levels of ceiling and floor effect which outperformed the proposed standards (Morton, 2008), with 0% and ~ 1% respectively.

3.4. Discussion

This review was concerned with the enhancement of functional status scales that specifically target community-dwelling older adults. It has been proposed that the relative standing of self-report ADL-IADLs could be enhanced by improving construct validities that are at least equivalent to those of physical performance measures. To address these challenges, this review chose to investigate constructs

related to scale hierarchy, ceiling effects, and establishing interval level measurement that enables the identification of construct under-representation.

Seven scales from this review were able to establish interval level measurement using parametric IRT procedures, thus enabling greater accuracy when considering change scores as well as identifying construct under-representation. With regard to construct under-representation all scales in this review presented with relatively large gaps in coverage, with the exception of McHorney and Coen (1990). When IRT methods are used to transform the ordinal nature of ADL scales to interval level data, diagnostic precision (McHorney, 1997), and sensitivity to clinical change are enhanced (Wirtz & Voigt-Radloff, 2008). Comparing disability measurements between items, between patients, or within patients between different moments in time is complicated. Change scores for Likert summative scores need to be interpreted with caution. It has been noted that assessing change in terms of estimated trait level rather than raw scores can yield more accurate estimates of change (May & Nicewander, 1998). If non-equal intervals exist between adjacent items, change scores for subjects with different levels of ability may misrepresent the amount of change, or fail to detect change in the latent trait (Morton, 2008). Furthermore, Fraley et al. (2000) demonstrated that analyses of change at the raw-score level and analysis of change using the latent-trait metric may lead to opposite conclusions. In one example, they displayed results showing that highly anxious individuals are relatively less stable over time when considered at the raw-score level, but more stable over time when considered at the latent-trait level. Thus, failing to understand the scaling properties of an instrument can lead to grossly inaccurate conclusions (Reise, Ainsworth, & Haviland, 2005).

Four scales met IRT standards for ascertaining item hierarchy at the individual level, as opposed to merely establishing item hierarchy at the population level. Despite the comprehensive coverage of McHorney and Cohen (1999), this manuscript made use of the 2PL IRT model which does not provide the added advantage of invariant item ordering; Ligtoet et al. (2010), point out that Sijtsma and Hemker (1998), proved that the graded response model used in McHorney and Coen (1999), does not imply

invariant item ordering. Invariant item ordering is clinically useful because improved understanding of the sequence of functional change or decline and its natural trajectory in aging would open up opportunities for thinking about early intervention and/or ways to change this trajectory (Daltroy, et al. 1992; Fieo, Watson, Deary, & Starr, 2010). Ligetvoet et al. (2010), reports that IIO is a strong requirement in measurement practice, and that researchers sometimes assume that fitting an IRT model implies that items have the same ordering by difficulty or popularity for all individuals, but this assumption requires modification. In following this rather strict criterion for IIO, our final pool of scales was relatively limited. This resulted in very few items that were common to other scales, thus allowing for only modest patterns of functional decline to emerge. That is, I had hoped to identify a dozen or so scales that exhibited invariant item ordering. In this way I would be able to make strong assertions as to a common hierarchy of functional decline that exists for community-dwelling older adults

It has been noted, within the last 25 years, that interest in measuring functional status among the nondisabled elderly has expanded dramatically because of the aging of the population and its implications for health care policy. As a result, measures of ADLs and IADLs have increasingly been applied to community-dwelling individuals, resulting in substantial ceiling effects (McHorney, & Tarlov, 1996). Four of the twelve scales were exceptional in reducing ceiling effects: Kempen and Suurmeijer (1990), reported 5% of subjects at the ceiling level; Fortinsky et al. (2003) also reported a ceiling effect of 5%; Haley et al. (2002), and Jette et al. (2002), observed a ~1% and 0% ceiling effect, respectively. However, it should be considered whether the success of the scales used in Kempen & Suurmeijer (1990), as well as Fortinsky et al. (2003), are being driven more by sample characteristics than scale sensitivity. Both scales were categorised in Table 1 as having the ‘least healthy’ samples of older adults. In the Kempen and Suurmeijer (1990) sample, all subjects were new users of professional home help, in addition to subjects being 77% female. Again, gender should be considered, as previous studies have reported gender differences in functional disability, with elderly women reported to have higher functional disability than elderly men (Allen, Mor, Raveis,

& Houts, 1993). The Fortinsky et al. (1990) sample were described as Medicare-eligible with a recent history of home care services, and one third of the sample was age 85 or above. Despite Haley et al. (2002) and Jette et al. (2002) also having a large proportion of female subjects, their subjects appear much healthier than the two other samples mentioned above. Thus, I can be more confident that the low percentage of ceiling effects has more to do with scale characteristics.

The success related to improved content validity can be attributed to the development of more difficult items. The items used in Haley et al. (2002) are very different than traditional IADL items (e.g., assessing the ability to 'Run a half mile'). In an effort to approach the novel status of a 0% ceiling effect, Haley et al. increased item difficulty. However, it has become apparent that 'newly developed' items designed to limit ceiling effects in high functioning populations lie outside the realm of daily experience, and thus may prove less reliable. For instance, questions about walking difficulty over a distance of one-quarter mile or more may be answered inaccurately simply because the respondent has not attempted to walk such a distance in quite some time (Simonsick, et al. 2008). Furthermore, it has been noted that the 'Vigorous activities' item (from a sample of chronically ill or psychiatric subjects) may have misfit due to lack of actual engagement in these activities within a typical day (Haley, McHorney, & Ware, 1994).

Lawton's instrumental activities of daily living (Wilms, et al. 2007) were thought to reflect a greater degree of complexity than the previously developed ADLs, and thus would be more applicable to a broader population of older adults. However, it seems that these traditional IADLs are most responsive to community dwelling older adults that show early signs of cognitive pathology, such as mild cognitive impairment. It has been shown that a majority of the traditional IADLs are more closely approximated with physical fitness than cognitive complexity (Ng, Niti, Chiam, & Kua, 2006). In an effort to reduce ceiling effects and to track change in community-dwelling older adults, scale developers have chosen to assess tasks that are more and more physically demanding, e.g., 'Run a half mile' or 'Vigorous activities'. However, the Late Life FDI scale presented in Jette et al. (2002), utilises difficult

items (as evidenced by a ceiling effect of 0%), while maintaining a degree of complexity, e.g., the ‘Travel out of town’ item or ‘Invite people into home’. And yet this scale does have two relatively large ‘gaps’ in coverage that might make tracking change over time problematic. Also these sorts of items may prove cumbersome for tracking progress in ‘prehabilitation’ (e.g., cognitive training) over relatively short intervention periods. It might be fruitful to explore the embedded components of a complex task such as ‘Travel out of town’, much the same way geriatricians have scrutinised the sub tasks involved in bathing (Gill, Han, & Allore, 2007; Naik, Concato & Gill, 2004).

Another avenue for increasing scale sensitivity in community-dwelling older adults is to alter the wording and thus the context in which activities are performed. Fries, Bruce, Bjorner, and Rose, (2006), provide a comprehensive review (with a mixed patient population) on the effects of altered context. In this review, Avlund et al. (1993), like Jette et al. (2002), explored atypical disability wording in an effort to reduce ceiling effects in community-dwelling populations (Avlund et al. (1993) is cited in the ‘close to inclusion’ section of this manuscript). Avlund et al. (1993) compared ‘tiredness’ and ‘reduced speed’ classifications, and found that the reduced speed scale was more effective in reducing ceiling effects. However, Avlund et al. (1996) advocated the rejection of the reduced speed scale (in favour of the ‘tiredness’) due to severe heterogeneity across age groups, as well as model fit difficulties. Avlund et al. (1996) also compared dependency (i.e., ‘do you need help?’) vs. tiredness and found that the tiredness scale was more suitable for measuring change among well older adults. At the same time, Fried et al. (1996) were altering scale classification by asking whether health or physical problems result in ADL-IADL tasks being completed with less frequency, or do such problems cause individuals to modify how they perform a particular functional task. Lastly, from this review, Schumacker (2004) used the uncommon categorization of ‘Do you have fear?’ performing various ADL-IADL activities. The result was massive floor effects, and the manuscript was ultimately excluded from this review because of poor reliability.

A primary advantage of IRT is the extension of reliability. Traditionally, reliability (i.e., the degree to which a scale is free of measurement error) has been used to assess a scale's *average* reliability. IRT however, is able to evaluate measurement error, or precision, at various stages along the scale continuum (e.g., disability construct). This is valuable because precision along the continuum is not uniform, and thus is expected to vary. This information is summarized with the information function, which allows for the estimation of the standard error of measurement for each subject's ability level. Despite the obvious utility, only one manuscript from this review chose to estimate the test information statistic—namely, Dubuc et al. (2004).

This review contains only one 2PL manuscript, which could be viewed as a study limitation. Some authors have suggested that 1PL models, as compared to 2PL models, are unsuitable as a final model for describing data resulting from functional status items (Martin, et al. 2007). Similarly, the fit of an IRT model can be examined with a likelihood ratio test, which assumes the more parameters that are used to describe item and subject behaviour, the better the model will fit the data (McHorney, 2002). However, the 1PL model is more robust (Martin, et al. 2007), and has the advantage of assuring that items can be ordered unambiguously, in the sense that their item characteristic curves do not cross (Spector & Fleishman, 1998); the 1PL procedure is the only well-known parametric IRT (as well as the rating scale model for polytomous items) model that has nonintersecting IRFs (Sijtsma & Molenaar, 2002). Additionally, fit statistics available for the 2PL model are barely reliable for scales containing few items, but are much more accurate in identifying items that don't fit model assumptions when large numbers of items are used (Cabrero-Garcia et al., 2005). A further limitation relates to the unavailability of data. This resulted in some logit data being extracted from figures rather than tables. This will merely have a small impact on the accuracy of reporting. Finally, several studies in this review use fewer than 100 subjects in their IRT analyses, which may be small even by Rasch standards. It has been proposed that a sample size of 100 will provide 95% confidence intervals for item calibrations. However, it has also been suggested that the adequacy of test targeting influences sample-size, and thus, a well-

targeted test may produce adequate location precision with less than 100 subjects (Morton, et al. 2008).

3.5. Conclusion

By improving construct validity and content validity for scales used to measure functional status researchers hope to improve accuracy and sensitivity: sensitivity to detect the early signs of functional decline in high-functioning older adults, and accuracy or precision to detect even small changes in ability. Several scales identified in this review were exceptional at reducing ceiling effects (e.g., Jette et al. 2002), or reducing gaps in coverage along the construct (e.g., McHorney & Cohen, 2002). However, a scale that exhibits both an exceptional reduction in ceiling effects, gaps in coverage, as well as IIO for community-dwelling older adults remains elusive. This review suggests that more work could be done to improve the validity of instruments used to assess functional status in community dwelling older adults.

Chapter four will mark the first empirical chapter of the thesis. In chapter four, I form several groups based on the most common chronic conditions and diseases experienced by older adults. These include coronary heart disease, diabetes, arthritis, comorbid arthritis and coronary heart disease, comorbid arthritis and diabetes, and comorbid arthritis and chronic obstructive lung disease. The ultimate aim was to determine whether these diverse conditions can be shown to adhere to a common hierarchy of functional decline, such that a generic disability instrument could be used for all group members.

Figure 3.1 Flow diagram for manuscript selection

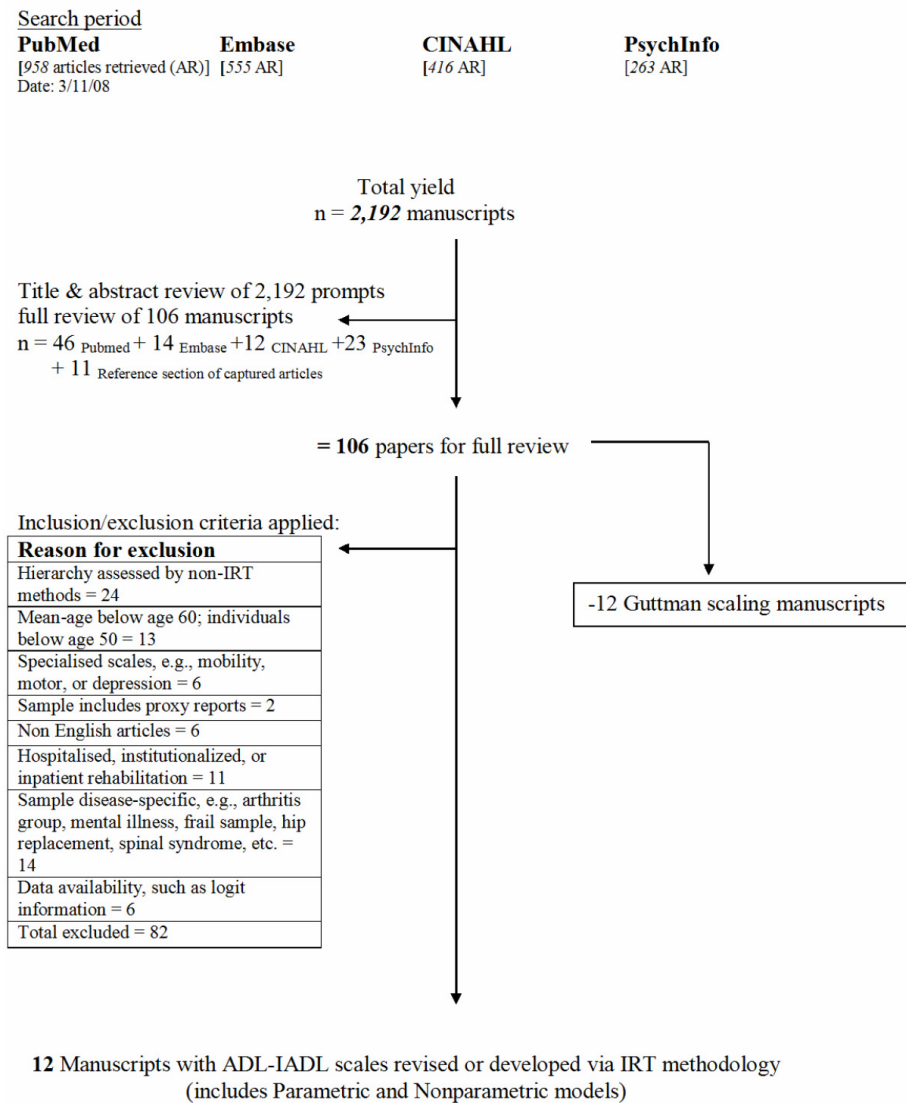


Table 3.1 Studies using IRT to establish a hierarchy of decline in ADL-IADL Scales

Study	ADL-IADL type	IRT model	# of items	Options	Sample studied	Reliability
Spector & Fleishman, 1998 (LH)	National Long-Term Care Survey ADL & IADLs	Rasch-model	15 ADL- IADL (1 item removed)	2(disabled vs. not disabled) #	Representative sample of disabled in the community *, Age 65 +, M = 79; n = 2,977	PS Reliability: .88
Haley et al., 2002	Late-Life FDI(function component)	Rasch-Rating Scale	27 ADL & IADL (5 items misfit)	5 (assessing difficulty)	Community-dwelling, Age 60-98, M 75.9, SD 8.5; n =150, 77% female	IS Index: 10.1
Sheehan et al., 2002	NHEFS disability questionnaire	Rasch-Partial Credit	24 ADL-IADL (1 item misfit)	4 (assessing difficulty)	Noninstitutionalized general population of older Americans, Age 57-86, M =62, n= 2,310	PS Index: 2.72 PS Reliability: .88
Jette et al, 2002	Late-Life FDI(disability component)	Rasch-Rating Scale	12 IADL (4 items misfit)	5 (assessing frequency)	Community-dwelling, Age 60-98, M 75.9, SD 8.5; n=150	IS Index: 9.39
Fortinsky et al., 2003 (LH)	Outcome and Assessment Information Set	Rasch-Partial Credit	15 ADL-IADL (zero items misfit)	3 to 6 (able to unable)	Community-dwelling, Medicare-eligible, with recent history of home care services, 1/3 of	Not reported
Dubuc et al., 2004	Physical Functioning Scale, PF -10	Rasch-Partial Credit	10 ADL-IADL (zero items misfit)	3 (limited by health)	Community-dwelling, n =75 , Age 60+, M 75.9, SD 8.5, 76% female	TIF: 4.5
Schumacker, 2004 (LH)	†	Rasch Partial Credit model	9 ADL- IADL (3 items removed)	2 (assessing fear)	Independent living facility (ILF), Age 65+, n=91	IS Index: 3.01 PS reliability: .64
McHorney & Cohen, 2000	†	2-Parametric Graded Response Model	166 ADL-IADL items derived through test equating	6 (difficulty)	Veterans Association sample with 75% being male, Age ≥ 65, n=3358	Not reported
Kempen & Suurmeijer 1990 (LH)	†	Mokken Scaling	18 ADL & IADL (zero item violations)	3 (difficulty)	Noninstitutionalized, Age 60 +,M =74.5 n =101, new users of prof. home help, 77% female	Rho coefficient.: 0.96
Kempen et al., 1995	Groningen Activity Restriction Scale (short)	Mokken Scaling	12 ADL-IADL	2 (difficulty)	182 residents of seniors' apartments, M=75, n=182	Rho coeff.: 0.87
Kempen et al., 1996	Groningen Activity Restriction Scale (GARS)	Mokken Scaling	18 ADL-IADL (zero item violations)	4 (difficulty)	Community-based sample, Age ≥ 57, n= 4773	Rho coeff.: 0.93
Watson et al., 2010	Townsend Functional Ability Scale	Mokken Scaling	6 items (3 violations)	3(difficulty)	Community-dwelling, All age 79, n=548	Rho coeff.: 77

Note:#= 'number of items', reflects the ending point, i.e., hierarchy confirmed after IRT application; M= Mean Age of sample; SD= standard deviation; LH = least healthy samples; PS index = person separation index; IS index = item separation index; reliability = person separation reliability; * =disabled defined as needing assistance with at least 1 ADL-IADL task; † scale type unspecified or 'newly devised'; ILFs are located near nursing homes &/or retirement homes

Table 3.2 Scales establishing interval level data

Spector & Fleishman, 1998	McHorney & Cohen, 2000	Sheehan et al., 2002	Jette et al., 2002
Shopping (-.83)	Scrub floor (1.75)	Heavy house chores (-2.49)	Active recreation (62)
Doing laundry (-.19)	Groceries 1 block (1.50)	Carry groceries (-1.70)	Volunteer job (53)
Bathing (-.10)	Iron clothes (1.25)	Walk two blocks (-1.48)	Travel out of town (53)
Mobility outside (-.02)	Stoop (1.00)	Light chores (-1.12)	Invite people to home (51)
Prepare meals (.29)	Cut toe-nails (.75)	Shop/run errands (-1.08)	Care for others (49)
Taking medicines (.38)	In/out of car (.50)	In/out bathtub (-1.02)	Visit friends & family (48)
Finances (.46)	Walk ½ block (.25)	Reach high, 5lb item (-.90)	Go out public places (47)
Mobility inside (.53)	Wash dishes by hand (.00)	Wash hair (-.22)	Care of home, inside (42)
Light housework (.56)	Balance checkbook (- 0.25)	Arise from chair (-.19)	Take care of errands (41)
Dressing (.60)	Go to the bank (- 0.50)	Pick up cloths (-.13)	Keep contact w/ others (36)
Transferring (.70)	Take vitamins (-0.75)	Up/down 2 steps + (-.12)	Take care of health (33)
Toileting (.94)	Wash face (-1.00)	Prepare own food (-.05)	Personal care needs (25)
Telephoning (1.1)	Answer telephone (-1.25)	In/out of car (.10)	
Incontinence help (1.60)	Drink from a glass (-1.50)	Dress self + tie shoes (.24)	
Feeding (1.61)		Wash & dry body (.33)	
		Open car doors (.45)	
		Cut meat (.48)	
		Open milk carton (.49)	
		Open jars (.56)	
		Write w/ pen/ pencil (.59)	
		Arise from bed (.75)	
		On/off toilet (.89)	
		Comb hair (1.17)	
		Turn faucets on/off (1.68)	
		Lift full cup/glass (2.75)	

Note: Brackets indicate large gaps in coverage, as a percentage of the total disability continuum; Numbers in parentheses represent logit intervals, with some scales making a further conversion to a 0-100 range for increased ease in interpretation

Table 3.2 Continued

Haley et al., 2002	Fortinsky et al., 2003	Dubuc et al., 2004
Run half mile (75)	Shopping (-3.35)	Vigorous activities (66)
Hike several miles (65)	Laundry (-3.34)	Walk 1 mile + (59)
Walk slippery surface (63)	Housekeeping (-2.61)	Up several flights (58)
Walk brisk mile (61)	Transport (-1.87)	Bend, kneel, stoop (57)
Run to catch bus (60)	Bathing (-1.15)	Walk several blocks (53)
Carry & climb stairs (59)	Prepare meals (-0.72)	Lift or carry groceries (52)
3 flights stairs inside (58)	Dress lower (-0.02)	Moderate activities (50)
1 flight outside (57)	Oral medication (-0.01)	Climb 1 flight (45)
Get up from floor (55)	Dress upper (0.56)	Walk 1 block (42)
Walk one mile (53)	Grooming (0.57)	Bath or dress self (32)
Walk several blocks (52)	Ambulation (1.64)	
Arise from low couch (51)	Telephone use (1.78)	
On/off a bus (49)	Toileting (2.01)	
Use step stool (48)	Transferring (2.78)	
Open heavy door (47)	Feeding (3.73)	
Up/down curb (46)		
Bend over (45)		
1 flight of stairs inside (44)		
Reach overhead (43)		
Make bed (42)		
Get in/out of car (41)		
Pick up chair (40)		
Walking inside house (37)		
On/off coat (35)		
On/off trousers (34)		
Wash dishes (33)		
Hold full glass (30)		

22%

21%

10%

13%

13%

29%

Table 3.3 Studies establishing invariant item ordering

Spector & Fleishman, 1998	Haley et al., 2002	Jette et al., 2002	Watson et al., 2010 *
Shopping (-.826)	Run half mile (75)	Active recreation (62)	Cut toe-nails (.72)
Doing laundry (-.188)	Hike several miles (65)	Volunteer job (53)	Up/down stairs (.30)
Bathing (-.103)	Walk slippery surface (63)	Travel out of town (53)	Get on a bus (.22)
Mobility outside (-.022)	Walk brisk mile (61)	Invite people to home (51)	Reach overhead shelf (.16)
Prepare meals (.294)	Run to catch bus (60)	Care for others (49)	Wash all over (.09)
Taking medicines (.380)	Carry & climb stairs (59)	Visit friends & family (48)	Tie knot in string (.04)
Finances (.460)	3 flights stairs inside (58)	Go out public places (47)	
Mobility inside (.528)	1 flight outside (57)	Care of home, inside (42)	
Light housework (.559)	Get up from floor (55)	Take care of errands (41)	
Dressing (.597)	Walk one mile (53)	Keep contact w/ others (36)	
Transferring (.699)	Walk several blocks (52)	Take care of health (33)	
Toileting (.944)	Arise from low couch (51)	Personal care needs (25)	
Telephoning (1.12)	On/off a bus (49)		
Incontinence help (1.60)	Use step stool (48)		
Feeding (1.61)	Open heavy door (47)		
	Up/down curb (46)		
	Bend over (45)		
	1 flight of stairs inside (44)		
	Reach overhead (43)		
	Make bed (42)		
	Get in/out of car (41)		
	Pick up chair (40)		
	Walking inside house (37)		
	On/off coat (35)		
	On/off trousers (34)		
	Wash dishes (33)		
	Hold full glass (30)		

Note: All scales present most difficult items first; * = scales assessed through nonparametric procedures; Numbers in parenthesis = logit values

4

A hierarchical decline of functional status across the most common chronic conditions/diseases experienced by older adults

4.1 Background

It is estimated that 20% of older U.S. adults have chronic disabilities (Manton & Gu, 2001). The disability pathway proposed by Nagi (1976) can be viewed as a theoretical construct, moving from disease to disability, as depicted in Figure 1.2 on page 33 (from chapter one). The utility of such a model relates to identifying the risk status of older adults; the Nagi framework can be used by gerontologists to understand the timing and opportunity for effective interventions by characterising their progression to severe disability and dependence. In Guralnik & Ferrucci (2003) the aim is to understand the pathway from 'Limitation to Disability'. In this chapter we are concerned with the pathway from 'Pathology to Impairment', and will attempt to elucidate whether various diseases and chronic conditions in older adults result in different patterns of disability.

For those aged 65 or older, as much as 88% of the population has at least one chronic condition (Hoffman, Rice, & Sung, 1996). This figure is rather alarming if we consider that chronic conditions are the most important determinants of disability (Valderrama-Gama, Damian, Ruigomerz, & Martin-Moreno 2002; Fried & Guralnik, 1997). In a large Canadian study of community-dwelling older adults aged 65 and over (n=9,008), five chronic conditions (foot problems, arthritis, cognitive impairment, heart problems, and vision) resulted in the largest risk of ADL-IADL disability. Taken as a whole, these conditions accounted for 66% of ADL-type disability and nearly 50% of IADL disability (Griffith, Raine, Wu, Zhu, & Stathothostas, 2010). Similarly, based on an average follow-up of 3 to 4 years, Wang (2002) found that selected medical conditions were associated with poorer functional

status, and that differences emerged in the patterns of associations between medical conditions and rates of functional decline. Coronary heart disease (CHD) was associated with increased rates of functional decline in all measures, i.e., ADL-IADL and performance-based physical function (PPF), whereas diabetes mellitus was associated with increased rates of functional decline in IADLs and PPF (Wang, 2002).

Despite its utility, the disablement process depicted in Figure 1.2 (page 33) can be considered too simplistic, as it is based on the notion of one disease causing one impairment causing one functional limitation causing one kind of disability (Avlund, 2004). It has been noted that a synergetic effect exists between the stages, which serves to increase the complexity of the disablement process, with multiple co-occurring diseases and impairments (Guralinik, 1994). Rantanen et al. (1999) showed that the odds of severe walking disability was ten times greater among those who had both strength and balance impairments compared with those who had only one or the other. Furthermore, Tinetti, Inouye, Gill, and Doucette (1995) examined how functional ability in older adults was influenced by four types of impairments (lower extremity impairment, upper extremity impairment, sensory impairment (vision, hearing) and affective impairment (anxiety or depression)). They found a significant increase in functional dependence as the number of these predisposing factors increased. Thus, Avlund (2004) notes that it might be appropriate to expand the model of the disablement process as shown in Figure 4.1 (page 97).

Understanding the ‘exchange’ between disease and disability is complex. This chapter aims to shed additional light on this exchange by emphasising the patterns of decline that exist within and between disease groups. We chose to examine the most common diseases and chronic conditions in older adults, which include co-morbid groups (Weiss & Ilowite, 2007).

4.1.1 Diseases and chronic conditions - Arthritis

Based on self-report measures, arthritis is the leading cause of disability in the United States (Centers for Disease Control and Prevention, 2001). Approximately 1 (37.6%) in 3 adults with arthritis reported limitation in their usual activities (Centers for Disease Control and Prevention, 2002). As the US population ages, the number of Americans ages 65 years with arthritis is projected to increase from 21.4 million in 2005 to 41.4 million by the year 2030 (Centers for Disease Control and Prevention, 2003). Arthritis was shown to double the risk of incident ADL disability (OR 2.2), and was the most prevalent chronic condition among this elderly population (57%) (Song & Chang, 2006). Cross-sectional studies indicate that elderly people with arthritis are much more likely to have limitations in mobility and ADLs. For instance, Kriegsman, van Eijk, Penninx, Deeg, & Boeke (1997) examined the relationship between seven chronic diseases and mobility performance and found that, adjusted for age and sex, all seven chronic diseases were significantly associated with a higher odds for mobility limitations, with odds ratios (ORs) varying from 1.38 for malignancies to 3.37 for arthritis. More recently, it has been found that middle-aged (Age 50 to 62) persons with arthritis were at higher risk for developing mobility and ADL difficulties that lead to loss of independence in late life (Covinsky et al., 2008). The details of this study are presented in Table 4.1 on page 98. Also, several longitudinal studies have shown that elderly people with arthritis are more likely to develop new disabilities (Leveille et al., 2001).

More sophisticated scaling procedures may serve to increase our understanding of the relationship between arthritis and the disablement process. Sheehan, Philpot, & Banerjee (2002) compared a national sample of civilian noninstitutionalized older adults, aged 50 to 77 (n=2310), with a sample of patients with definite classical rheumatoid arthritis (n=605). Sheehan et al. employed the Rasch partial credit model (Andrich, 1978) to confirm a hierarchy of functional decline. This was a parametric item response theory model capable of enhancing scale accuracy by establishing interval logit units. Sheehan et al. assessed functional status with 19 ADL-IADL items taken from the NHANES Follow-up Study (NHEFS), and the Health Assessment Questionnaire (Fries, Spitz, Kraines, & Holman, 1980). Performance on

the scales varied considerably between the two groups, with the largest change in item hierarchy being observed for ‘Open milk’ container and getting ‘In and out of a car’. From a clinical or theoretical standpoint the former observation (i.e., Open milk) seems relatively obvious, as rheumatoid arthritis most commonly affects small joints of the hands, feet and cervical spine. However, the fact that getting in and out of a car was the 4th easiest task of 19 items, compared with the 11th easiest task for the healthy sample may be counterintuitive; Rheumatoid arthritis (RA) often affects symmetrical joint involvement of the knees, which might impede one’s ability to move in and out of a vehicle. An examination of Fries et al. (2006) may serve to clarify the above observation. Fries et al. conducted a hierarchical confirmatory factor analysis on a large sample of RA patients and found that a four factor model of physical function items provided a better fit than a one, two, or three factor model. The factors were titled Upper Extremity, Lower Extremity, Central: neck and back (trunk), and Compound Activities. Fries et al. noted that these clusters were in agreement with the PROMIS postulated sub domains of ‘physical function’. The Fries et al. study is relevant to the Rasch hierarchy observed in Sheehan et al. (2002), in that the relative ease for RA subjects in getting in and out of a car has less to do with knee torque or knee joint loading and more to do with the neck and back body regions that may be spared the negative effects of RA. This position is supported by another fact: standing up from a chair and getting out of bed were cited as tasks in the ‘Central: neck and back’ factor, and also reflected items in the Sheehan et al. study that were less difficult for RA subjects than normal or healthy subjects, from a hierarchical perspective.

4.1.2 Diabetes

The prevalence of diabetes increases with age, reaching a plateau at 10–20% (depending on study populations, screening strategies and diagnostic criteria) in people over 70 years old (Wild, Roglic, Green, Sicree, & King, 2004). Patients with diabetes are two to three times more likely to report disability than those without diabetes (Tucker, Bermudez, & Gastaneda, 2000). It has been asserted that declines in functional status associated with the presence of diabetes have been examined less frequently, despite the consequences to independence and quality of life (Sinclair et

al., 2008). It is clear that diabetes results in substantial personal health burden (Sinclair, Gray, Lunec, & Barnett, 1993), but the underlying mechanisms that drive the diabetes/disability relationship is less clear. Some research points to factors such as neuropathy and vasculopathy (Tesfaye et al., 2005), and the presence of inflammatory markers (such as C-reactive protein and interleukin-6) probably poses additional independent risk of functional limitation in the diabetic populations (Figaro et al., 2006). Figaro et al. (2006) investigated C-reactive protein's relationship to mobility performance. Figaro et al. found, after adjusting for several demographic and clinical factors (age, race, sex, performance score at baseline, use of anti inflammatory drugs, current smoking status, PVD, heart failure, CHD, body mass index, current oestrogen use, and statin use), that those in the highest tertiles of C-reactive protein and interleukin-6 had the highest risks of functional limitation: HR 1.5 (95% CI, 1.2–1.6) and 1.7 (1.4 –2.0), respectively. Other explanatory factors include whether disability is largely the result of hyperglycaemia or of chronic complications of the disease (Bruce, Davis, & Davis, 2005). The question is an important one, because therapies targeting hyperglycaemia may improve disability in diabetic persons. Bossoni, Mazziotti, Gazzaruso, and Martinelli (2008) investigated 'controlled' diabetes by regulating serum glycosylated haemoglobin (HbA1c). Those individuals who did not meet glycaemic targets (i.e., controlled diabetes) presented with IADL disability scores that were significantly higher than diabetic patients with well-controlled disease. Bossoni et al. speculated that IADL disability may be a consequence rather than a cause of uncontrolled hyperglycaemia. Furthermore, it was found that even mild increases in plasma glucose levels may be sufficient to determine disability in older adults, and IADLs may be impaired early during the natural history of type-2 diabetes. A potential mechanism underlying hyperglycaemia-induced IADL disability may be abnormal muscular performance.

Muscle weakness in diabetes has been considered a rare manifestation associated with severe diabetic neuropathy (Dyck et al. 1993). However recent studies, with the use of quantitative assessments of muscular function, have shown that skeletal muscle strength, particularly in the lower extremity region, is typically lower in adults with diabetes than in non-diabetics. Park et al. (2007) observed that older

adults with type 2 diabetes lost 13.5% of their knee extensor strength, whereas those without diabetes lost 9.0% of initial strength in 3 years. Surprisingly, a 50% more rapid decline in the knee extensor strength in older adults with diabetes was not accounted for by a greater loss of leg muscle mass. Thus, muscle quality of the lower extremities (i.e., functional impairment), and not muscle mass may lead to mobility impairment (Park et al., 2007). In fact, there is some evidence that hyperglycaemia can affect contractile function and force generation in muscles (Anderson, Randles, & Kendall, 2004).

The disability/diabetes relationship is further complicated by the fact diabetes shares many of the determinants associated with disability. Vereberege and Jette's (1994) list of disability determinants includes: social and economic factors, lifestyle-related risk factors (e.g. poor nutrition, obesity, smoking, sedentary preference), psychological factors (personality type, coping strategies), psychiatric conditions (mood disorders, changes in cognition) and a range of disabling medical conditions (arthritis, cardio-pulmonary disease, cancer, & stroke). Bruce, Markovic, and Fonarow (2011) note that the above indicators are also implicated in the development of diabetes.

4.1.3 Chronic obstructive pulmonary disease/chronic obstructive lung disease

The prevalence of *chronic obstructive pulmonary disease* (COPD) in individuals between 40 and 80 years of age is 10.2% (Spain) and increases with age, tobacco consumption and lower educational levels (Miravittles & Anzueto, 2009). Furthermore, both diagnosed and undiagnosed obstructive lung disease are known to be associated with increased functional limitation (Coultais et al., 2001). Eisner et al. (2008) reports that COPD affects a multitude of body systems remote from the lung, including a broad array of physical functional limitations (e.g., including lower extremity functioning, exercise performance, skeletal muscle strength). Table 4.2 (page 99) from Eisner et al. compares healthy controls to those with COPD, and reported an odds ratio of 7.8 (95%CI 4.0 to 15.1) for developing difficulties in moderate activities (e.g., pushing a vacuum cleaner) and an odds ratio of 13.2 (8.0 to 22) for self-reported limitation in walking up stairs.

Similar findings were reported by Mannino, Homa, Akinbami, Ford, and Redd (2002) for those with severe COPD: after adjusting for age, sex, race, smoking status, body mass index and the presence of chronic disease they reported an Odds Ratio of 8.4 (3.6, 19.9) for 'unable to walk ¼ of a mile and OR 16.2 (4.9, 53.5) for needing help with daily activities.

4.1.4 Coronary heart disease

A population based study investigated the association of disability with Coronary Heart Disease (CHD) incidence in older adults (age 71 +) (Corti, Donovan, & Holman, 1996). This study assessed mobility disability and mobility plus ADL disability and found that both measures, within a 4-year period, coincided with twice the risk of CHD mortality. Furthermore, there was a gender difference between both measures of disability and risk of nonfatal CHD, with women *only* presenting a 50% greater risk than men. A more recent study, Plichart et al., (2010), appears to replicate these findings by demonstrating an association between disability severity (four categories *progressing* from no difficulty in mobility, IADLs, or ADLs to difficulty in all three domains) and fatal CHD. This was a longitudinal study with a six-year follow-up period. Because disability might be associated with prevalent CHD, Plichart et al. began by excluding 18% of the sample (n=1,626) who had a past history of CHD. After 5.2 years 264 subjects experienced 264 first coronary events (includes 55 fatal events). After adjusting for CHD risk factors, those with difficulty in both mobility tasks and IADL tasks were 1.7 times (95% confidence interval (CI)=1.0–2.7) greater risk of overall CHD than subjects with no mobility, IADL, or ADL difficulty. The subjects with fatal CHD presented with a hazard ratio of 3.5, 95% CI=1.3–9.3; *P* for trend=.01). The authors conclude that disabled community-dwelling older adults are at substantially greater risk of fatal CHD and that prevention of CHD in that population is therefore of particular importance.

Kattainen et al (2004) examined disability as a predictor of mortality among 4,501 men and women aged 45 and over in a national sample of the Finnish population. Disability was defined as having *marked* difficulty in one or more ADL-IADL

tasks, thus covering a wide range with regard to severity (e.g., moving about in the house or dressing and undressing vs. walking 400m or carrying a 5-kg shopping bag). The authors observed that difficulty in ADL-IADL measures is associated with a substantial increase in mortality from CHD and all causes in middle-aged and younger elderly men with CHD, however significance did not hold for women. Kattainen et al. also found that disability (ADL-IADL difficulty) was associated with mortality risk in both men and women who presented with no CHD at baseline. The authors concluded that disability has an independent effect over and above the contribution of CHD risk factors.

4.1.5 Comorbidity

The formal definition of comorbidity is the concurrent presence of two or more medically diagnosed diseases in the same individual (Fried, Ferrucci, Darer, Williamson, & Anderson, 2004). Wieland (2005) adds that the one or more conditions should be accompanied by a primary or “index” disease. In 2000, an estimated 57 million Americans had multiple chronic conditions, and the number is projected to increase to 81 million by 2020 (Wu & Green, 2000). The rate of comorbidity is thought to increase as we age, with 69% of persons over age 65 having 2 or more chronic illnesses (Hoffman, Rice, & Sung, 1996). A major contributor to adverse health in ageing relates to the presence of comorbid conditions, as comorbidity increases the risk of disability and mortality increases (Fried et al., 2004). It has been shown that a higher number chronic diseases or conditions is routinely associated with the increased prevalence of mobility limitations, both in cross-sectional (Guralnik et al., 1990) and longitudinal studies (Guralnik et al., 1993).

Comorbid conditions are more difficult to treat: they are more complex in terms of competing risks and potentially incompatible therapies (Fried, 2003). The difficulty in completing a particular functional task caused from one condition may be influenced by the second condition in a comorbid state (Fried et al., 1999). For example, the influence of cardiovascular disease on functional status becomes less clear in those individuals diagnosed with CVD and COPD because COPD can act as

a predictor of CVD hospitalization and mortality (Sidney, 2005). Fried et al. (2005) points out that comorbidity may have a synergistic effect (depicted in Figure 3). For example, Ettinger, Davis, Neuhaus, and Mallon (2004), observed that, for the development of mobility disability even the sum of the risk posed by heart disease (odds ratio (OR) = 2.3) and the risk posed by osteoarthritis (OR = 4.3) is considerably less than the risk posed by the combination of the two (OR = 13.6). The complexity of comorbidity is highlighted by the fact that the simple summation or accumulation of morbidity does not necessarily translate into greater disability. For example, (Marchionni et al., 1996) has shown that cardiac failure (CF) and comorbid chronic obstructive pulmonary disease, hearing impairment, gastrointestinal tract disease, or osteoarthritis did not increase the prevalence of disability beyond the independent effect of CF. However, when CF coupled was coupled with visual impairment, previous stroke, or urinary incontinence the effect on disability was much larger. Lastly, Wieland (2005) asks the question: what is the role of convergent mechanisms? For instance, does heart disease lead to mobility disability through the loss of exercise tolerance? In the preceding section of this chapter, it was observed that mobility disability acts to increase the risk of heart disease. Wieland suggests that the answer to his question is still unclear; Wieland conveys, aside from evidence that some particular conditions are more likely to co-occur than others, we lack clear answers to such basic questions. It was further suggested that an improved understanding of comorbidity or multi morbidity may be gained by enhancing our definition of severity, that is, severity as it relates to acuity, chronicity, or perhaps progression.

As further evidence for the complexity inherent in multimorbid states, Wieland (2005) asks, how do we study the relationships between multimorbid conditions and functional impairments, when functional impairments are among the criteria for defining some of these conditions. Due to the complexity of multimorbid conditions and the fact that the rate of such conditions is likely to rise as the number of cohorts in the oldest-old range increases, greater clarity of the disablement process is essential to maintaining a reasonable quality of life for older adults. Fried conveys “Improved mapping of the relationship between individual or multiple diseases and

specific types of functional limitations will lead to better understanding of the risk factors for disability, and increase the likelihood of developing effective interventions to prevent or minimize disability' (Fried, 1994, p.758).

The analyses in this chapter are concerned with mapping the relationship between common chronic conditions, as well as comorbid conditions, and mobility disability. This will be achieved by enhancing the interpretive power of the instrument (functional status scale) used to measure mobility performance in older adults. For instance, I will seek formally to confirm a hierarchy of functional decline. It has been suggested that establishing an item hierarchy (based on increasing levels of difficulty) enhances a scale's internal validity: when the items demonstrate statistical goodness of fit to item response theory (IRT) models, the scale can be said to be unidimensional. IRT unidimensionality indicates that the items on the scale belong to a single line on which items range from those that are easily performed to those that are hard to perform, thus supporting the scale's internal validity (Bond & Fox, 2001). Additionally, sequential functional loss scales tell more than the typical simple summation of functional loss, and may have predictive value to the clinician monitoring an elderly patient. If the sequence is accelerated or out of the usual order, such as seen in patients with arthritis, it may indicate the need for interventions (Daltroy et al., 1992). Finally, examination of the sequence of loss may help characterize adaptations to impairment and differences among subgroups.

4.2 Method

4.2.1 Sample

The Cardiovascular Health Study (CHS) is a prospective, observational study designed to determine the risk factors for and consequences of cardiovascular disease in older adults. In 1989 and 1990, a total of 5201 men and women aged 65 years or older were recruited in four US communities to participate in CHS: Sacramento County, California; Washington County, Maryland; Forsyth County, North Carolina; and Allegheny County (Pittsburgh), Pennsylvania. Potential participants were identified from a random sample, stratified by age group (65-74, 75-84, ≥85 years)

from the Health Care Financing Administration (HCFA) Medicare Enrolment Lists. In 1991 this cohort was increased to 5888 subjects with the addition of minorities into the sample. All persons thus identified and age-eligible household members who were planning to reside in the community for at least 3 years were eligible to participate. Exclusion criteria included being wheelchair bound in the home, unable to participate in the examination at the field center, or undergoing active treatment for cancer. Fifty-seven percent of eligible subjects agreed to participate. Despite participants reporting an average of two medications used and three chronic diseases or conditions, 56% of the sample reported no difficulty with the 17 tasks used to assess functional status related to daily activities (Fried et al., 1994). The full details of the sampling and recruitment have been published previously (Fried et al., 1991).

The evaluation consisted of a home interview and then a clinic examination. The interview assessed self-report of physical function (see below), demographic characteristics, cognition, psychosocial factors, and medical history. Depression was assessed using the Center for Epidemiologic Studies Depression (CES-D) Scale (Orme, Reis, & Herz, 1986). Medical history of non-cardiovascular diseases was obtained through self-report of physician-diagnosed conditions. All participants had annual clinic examinations up to 1998–1999, including the Modified Mini-Mental State Examination (Teng & Chui, 1987) and the Digit Symbol Substitution Test (Wechsler, 1981). To be eligible for participation in the study, sampled individuals or their household members had to be 65 years of age or older and noninstitutionalized. Potentially eligible persons were excluded if they were wheel-chair bound in the home; were unable to participate in the field center (clinic) examination; were receiving treatment for cancer, including radiation or chemotherapy, or hospice care; or were participating in another research study.

Our investigation begins with the selection of the most common diseases and chronic conditions for adults age 65 and older, as outlined by Weiss et al. (2007). Examining a representative sample of US noninstitutionalized civilians aged 65 years or older (n=4349) Weiss et al. observed that arthritis was the most common chronic condition/disease, followed by CHD, co-occurrence of CHD & arthritis, COPD & arthritis, diabetes & arthritis, diabetes, and COPD. Due to the presence of large

ceiling effects (only 21% of the total sample endorsed difficulty with the most difficult mobility task, and just 2% of the sample reported difficulty with the most challenging self-care task), the present investigation was limited to those subjects aged 75 to 84, which resulted in higher levels of reported difficulty. This reduced the sample to $n=1749$. I proceeded to exclude all subjects with Low BMI (< 18.5), Current smokers, MMSE (≤ 24), Eye disease, and Depression (score ≥ 16 from the Center for Epidemiological Studies Depression Scale) (Orme, Reis, & Herz, 1986). These subjects were excluded to obtain six groups (arthritis only, CHD only, diabetes only, co-occurrence of CHD & arthritis, COPD & arthritis, diabetes & arthritis) that were relatively homogeneous, and mutually exclusive. This reduced the sample to $n=1372$, with the following subgroup distribution: arthritis only ($n=485$), CHD only (120), diabetes only (192), co-occurrence of CHD & arthritis (161), COPD & arthritis (228), diabetes & arthritis (186).

4.2.2 Assessment of functional status

An interviewer-administered questionnaire ascertained self-reports of the difficulty in performing 17 specific tasks of daily life (see appendix, Table A1, for items and their factor loadings), using a modified version of the Health Interview Survey Supplement on Aging Questionnaire (Fitti & Kovar, 1987). This asked for the participant's assessment of whether he or she had difficulty with bathing, dressing, eating, using the toilet, walking around their own home, and getting out of a bed or chair (i.e. ADLs); heavy housework, light housework, shopping for personal items, preparing their own meals, managing their money (such as paying bills), or using the telephone (i.e. IADLs); with lifting or carrying something as heavy as 10 lb, walking 1/2 mile, walking up 10 steps (i.e. Rosow-Breslau scale); or with reaching out with their arms, or gripping with their hands. Difficulty was ascertained with a three categories: 'some difficulty', 'a lot of difficulty', or 'unable to do'.

Previous examination (Fried et al., 1994) of the CHS sample, using principal components analysis and orthogonal rotation, identified four factors from the self-report of functional status. The goal of the Fried et al. study was to elucidate the relationship between various performance measures and the modified Health

Interview Survey Supplement on Aging Questionnaire. The four factors were labelled Mobility, Complex (IADL), Self-care (ADL), and Upper Extremity. Due to massive ceiling effects (noted below) in the IADL and ADL factors, I chose to limit my investigation to the Mobility factor. The proportion of subjects who reported 'disabled' or 'unable to do' for the IADL and ADL factors was as follows: Pay bills (1%), Meal preparation (.07%), Shopping (2.3%), Light housework (.07%), Telephone use (.03%), Using toilet (.01%), Dressing (.02%), Bathing (.05%), and Eating (.01%). Lastly, an examination of item hierarchy proved untenable for the Upper Extremity factor, as it contained only two items—Gripping and Reaching.

We employed the Mokken scaling procedure to determine whether a common hierarchy of functional status (i.e. Mobility factor) could be identified among the six most common chronic conditions associated with older adults. The Mobility factor included the following items: 'Do you have difficulty walking one-half a mile, about 5-6 blocks?'; 'Do you have any difficulty walking around your home?'; Do you have any difficulty getting out of a chair or bed?'; Do you have any difficulty walking up 10 steps?'; Do you have any difficulty with heavy housework, like scrubbing floors or washing windows or yard work, like raking leaves, or mowing?'; Do you have difficulty lifting or carrying something as heavy as 10 pounds, such as a bag of groceries?'.

4.2.3 Statistical analyses

The mobility items from the modified Health Interview Survey Supplement on Aging Questionnaire were examined using the Mokken Scaling Procedure (version 5.0 for windows) (Molenaar & Sijtsma, 2000). This is a nonparametric hierarchical scaling method based on the principles of item response theory (IRT), which was employed to confirm a hierarchy of functional status, and to determine the discriminatory power of each item in the scale. The latter objective will inform researchers as to which items are most closely related to the clinical dimension of interest, and the former will allow for the identification of varying levels of item or task severity (i.e., confirm item hierarchy). A scale that adheres to a hierarchy will have items that can be ordered by degree of difficulty along the trait being investigated.

The Mokken procedure can be separated into two levels of analyses: scalability and invariant item ordering (item hierarchy). The scalability component assesses the degree to which patients can be accurately ordered on the latent trait by means of their sum score (Roorda, Scholtes, Van der Lee, Becher, & Dallmeijer, 2010). Scalability is investigated by means of the monotone homogeneity model (MH). For the MH model to hold for a scale three common assumptions in IRT must be met: unidimensionality, local independence, and monotonicity. Several statistics are produced to confirm that the MH holds. Scale criteria are met when the coefficients of scalability for all item pairs (H_{ij}) are positive, whereas the scalability coefficients for the items in relation to the scale at issue (H_i) and for the scale (H) are at least .30. Higher values for H_i and H imply fewer violations and thus a better scale. A rule of thumb is that a scale is considered to be weak if $.30 \leq H < .40$. Medium scalability is obtained if $.40 \leq H < .50$. A scale is considered to be strong if H is equal to or greater than .50 (Van der Ark, 2007). MH allows for the ordering of persons on the latent trait by the sum of the item score – an essential requirement for a psychological test (Watson & Deary 2010). The Mokken procedure also produces a Rho coefficient, which is a reliability statistic concerned with establishing the repeatability of the sum score. Most theorists agree that a Rho over .80 is desirable, and a Rho over .70 is a minimum requirement (Kempen & Suurmeijer, 1990).

As stated above, the Mokken model also provides a diagnostic for confirming invariant item ordering (IIO). The IIO property is crucial for establishing hierarchical scales. The order of the items in terms of difficulty should be the same for all respondents whatever their latent trait value (Sijtsma & Junker, 1996). Establishing a formal hierarchy of decline, or invariant item ordering (IIO), should enhance construct validity. Ligtoet, Van der Ark, Te Marvelde, and Sijtsma (2010) convey that IIO is strong requirement in psychometrics, and that researchers wrongly assume that fitting any IRT model implies that the items have the same ordering by difficulty for all subjects. The IIO property goes beyond merely ordering subjects by popularity or mean score. Sijtsma, Meijer, and van der Ark (2010) indicate that it would be an aggregation error to infer from mean item ordering that item (group level) holds for

each individual respondent. The criteria for IIO are met when the percentage of negative coefficients at the level of the individual subjects (H_a^T) is less than 10% and the coefficient for total set of subjects (H^T) is at least .30 (Roorda et al., 2010).

4.3 Results

Two of the groups analysed were unable to meet the assumptions relating to the monotone homogeneity model. More specifically, the sign for all pairs of items must be nonnegative and items with small positive H_i (below 0.3) should be excluded because they contribute very little to reliable person ordering (Sijtsma & Molenaar, 2002). Within the ‘Diabetes Only’ group, Item 15 (carry groceries) had a negative covariance with item 2 (walk in home). The ‘carry groceries’ item also proved problematic for the ‘Arthritis & CHD’ group. However, in this group ‘carry groceries’ was excluded due to a lower bound H_i value below .30. Because item 15 presented with difficulty in two of our groups I removed this item and re-ran the analysis with five items rather than six, so as to press on with our aim of identifying a common mobility hierarchy for all six groups. This five-item hierarchy for each group is presented in the six tables below (Tables 4.3, a-f on pages 100-101).

Tables 4.3a to 4.3f include estimates of Loevinger’s H coefficient. Mokken Scaling violations are defined as the deviation of the observed data structure from the perfect scalogram structure, i.e., an easier item wrong, or a more difficult item right (Sijtsma & Molenaar, 2002). The set H coefficient is used to measure the extent to which persons always appear in the same relative order and validates their use in forming a unidimensional latent variable (Watson, Deary, & Austin, 2007), with a higher value H coefficient reflecting fewer violations. A minimum H of .30 is recommended, but one can speak of a “strong” scale for values $\geq .50$ (Kempen et al., 1995). Tables a (‘CHD Only’) and c (‘Diabetes Only’) present with the largest H total, .76 and .74 respectively, and well above the threshold for a strong scale. The remaining tables, b, d, e, and f all present H total coefficients above .50, reflecting strong scales. Table c (Diabetes Only), also presents with the highest reliability coefficient, Rho of .84. The remaining tables present reliability coefficients just above the minimum requirement

of .70, with a range of .71 to .76. All scales surpassed the minimum requirements ($H^T \geq .30$ and $H_a^T < 10\%$) necessary to establish invariant item ordering (IIO) or item hierarchy, with the exception of ‘Arthritis & COPD’ presented in table f. As seen in table f, an H_a^T 10.5% lies just above the threshold required for establishing an item hierarchy of functional decline. Two scales, presented in tables c and e, exhibit differential item functioning (DIF), ‘Diabetes Only’ and ‘Diabetes & Arthritis’. DIF, or item bias, draws attention to the issue of making valid comparisons between subgroups of patients (Roorda et al., 2008). Both of these scales exhibit a reversal in the most difficult items. That is, unlike the four remaining patient groups that exhibit a difficulty hierarchy that begins with Heavy Housework and progresses to difficulty with Walking ½ Mile, in the two diabetes groups Walking ½ Mile is the most challenging item.

4.4 Discussion

This chapter sought to establish whether the most common chronic diseases/conditions experienced by older adults follow a common hierarchy of mobility decline. The ability to confirm a common hierarchy of functional status for a diverse set of chronic conditions commonly experienced by older adults should prove beneficial to gerontological researchers. Arguably, the most fundamental parameter in scale analysis is the difficulty parameter (Wilson, 2005), which allows item/task responses to be linked with the construct under investigation. That is, the difficulty parameter allows the researcher to define whether the subject exhibits more or less of the construct, e.g., one’s level of functional reserve.

The initial mobility scale consisted of six functional status tasks. The early stages of analyses warranted the removal of one item (carry groceries), due in part, because it violated assumptions of unidimensionality. The removal of this item was not entirely unexpected, as it consistently demonstrated a low H_i statistic. Loosely speaking the H_i coefficient expresses the extent to which item i fits together with the other $k-1$ items of a scale (Sijtsma & Molenaar, 2002). Factor analysis prompted Fried et al. (1994) to place ‘carry groceries’ and ‘heavy housework’ within the Mobility factor.

However, Fried et al. noted that both of these items also showed a relationship to the Upper Extremity factor. In all six patient groups these items presented with the lowest H_i values, suggesting that these tasks contained the least amount of variance associated with mobility performance. Despite the Mokken procedure meeting the assumption of unidimensionality, minor abilities can still influence response patterns. It has previously been asserted that responses to a set of items are multiple determined, such that multiple minor abilities are required to respond to items (Nandakumar, 1994).

The analyses here established a formal hierarchy of mobility decline for five of the six conditions examined. The 'COPD & Arthritis' group was the only mobility scale that did not adhere to IIO. Eisner et al. (2008) reports that COPD affects a multitude of body systems remote from the lung, including a broad array of physical limitations. This fact may make it more difficult to order items and or subjects along the construct under investigation. It should be noted that one of the items, heavy housework, presented with the lowest H_i coefficient for any item in all six scales. The removal of this item is likely to result in establishing IIO, particularly since one of the two criteria used to assess IIO was met.

Two scales exhibit differential item functioning (DIF), 'Diabetes Only' and 'Diabetes & Arthritis'. DIF is concerned with the question as to whether or not the likelihood of item endorsement is equal across subgroups. It is noteworthy that the two subgroups that exhibited DIF were comprised of diabetes subjects. In both scales the most difficult item was 'walking ½ mile', rather than 'heavy housework'. This result may be explained by the fact that skeletal muscle strength, particularly in the lower extremity region, is typically lower in adults with diabetes than in non-diabetics (Park et al., 2007).

The difficulties encountered (i.e., DIF in two scales and an inability to confirm IIO in one scale) in establishing a common hierarchy of functional decline for all six patient groups involves the 'heavy housework' item. The removal of this item will likely result in confirming IIO for all subgroups, and the elimination of DIF. However,

reducing our scales to just four items may serve to decrease the reliability of sum scores. Furthermore, the removal of this item would increase ceiling effects in a rather dramatic fashion. Ideally, it would be best to replace the ‘heavy housework’ item with a challenging task that more closely approximates mobility difficulty.

4.5 Conclusion

The results of this chapter demonstrate the potential for mobility reserve to be used as an early indicator (or preclinical marker) for ADL disability in community-dwelling older adults. More importantly, Mokken analyses indicate that a common hierarchy of functional decline was observed for a diverse set of conditions and diseases that are common among community-dwelling older adults. Such an indicator could be used to identify functional status declines relating to severity in diverse patient populations.

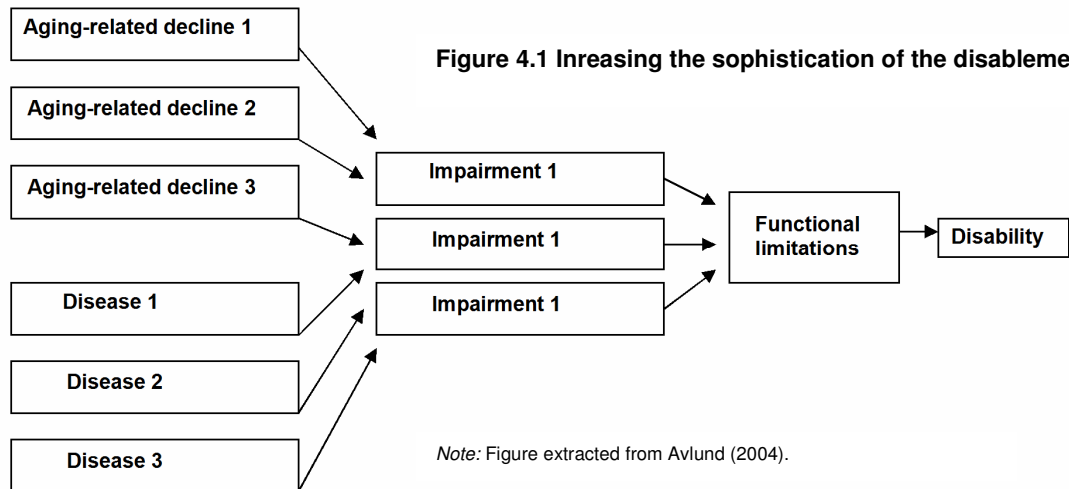


Table 4.1
Effect of Arthritis Compared with That of Other Chronic Conditions on the Development of Mobility or Activity of Daily Living (ADL) Difficulty

Group	Mobility/ADL Difficulty, %	Hazard Ratio (95% Confidence Interval)	
		Unadjusted	Adjusted*
No arthritis or other chronic condition (n = 2,658) [‡]	11	1.00	1.00
Arthritis, no other chronic condition (n = 853)	23	2.10 (1.66–2.66)	1.91 (1.59–2.44)
No arthritis, at least one other chronic condition present (n = 2,683)	24	2.30 (2.00–2.66)	1.74 (1.46–2.08)
Both arthritis and at least one other chronic condition (n = 1,349)	43	4.76 (4.14–5.46)	2.79 (2.31–3.38)

* Adjusted for age, sex, race, education <12 years, employed <20 h/wk, income, net worth, body mass index, exercise ≥3 times per week, difficulty jogging 1 mile, difficulty climbing one flight of stairs.[†]Other chronic conditions were hypertension, diabetes mellitus, cancer, chronic lung disease, myocardial infarction, congestive heart failure, stroke, and obesity. (Table from Covinsky et al., 2008).

Table 4.2
Impact of COPD on Self-reported Functional Limitation

	ALL SUBJECTS	GOLD Stage II+	GOLD II+ CONTROLLING FOR COMORBIDITIES
Functional limitation measure	Odds ratio for COPD vs. referents (95% CI)	Odds ratio for COPD vs. referents (95% CI)	Odds ratio for COPD vs. referents (95% CI)
Self-reported functional limitation [‡]	6.4 (3.7 to 10.9) P<0.0001	7.0 (3.9 to 12.3) P<0.0001	6.0 (3.4 to 10.7) P<0.0001
Limitation in moderate activities ^{‡‡}	7.6 (4.0 to 14.4) P<0.0001	8.7 (4.5 to 16.9) P<0.0001	7.8 (4.0 to 15.1) P<0.0001
Limitation in climbing several flights of stairs	11.7 (7.3 to 18.6) P<0.0001	14.0 (8.5 to 23) P<0.0001	13.2 (8.0 to 22) P<0.0001

Multivariate logistic regression analysis controlling for age, sex, race, height, smoking, education [‡]“Unable to do” or “severe limitation” in a battery of basic physical activities (see Methods) ^{‡‡}Health limits “a lot” in performing moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf. The fourth column shows results for more severe COPD only (GOLD Stage II or greater) vs. referent group (n=742 COPD cases vs. 302 referent subjects). The fifth column shows the same group with additional statistical control for cardiovascular comorbidities and diabetes. (Table from Eisner et al., 2008).

Table 4.3a Mokken Scaling Procedure Applied to 'CHD Only' Group

Item	Mean	ItemH	Z
Heavy housework	0.19	0.61	6.56
Walk 1/2 mile	0.11	0.69	7.16
Climb 10 steps	0.04	0.90	8.30
Rising from a chair	0.02	1.00	8.19
Walk around home	0.01	1.00	6.56
<i>H</i> total: .76, <i>H</i> ^T :.63, <i>H</i> _a ^T :4.2%			
Rho = .76			n=106

MSP = Mokken Scaling Procedure; Range for each item is zero or one (zero reflects no impairment & one indicates difficulty). Total H represents the H coefficient for the entire scale. H^T H_a^T = HTrans coefficients. Rho=Mokken reliability statistic.

Table 4.3b Mokken Scaling Procedure Applied to 'Arthritis Only' Group

Item	Mean	ItemH	Z
Heavy housework	0.21	0.52	13.20
Walk 1/2 mile	0.14	0.52	15.27
Climb 10 steps	0.11	0.48	14.17
Rising from a chair	0.04	0.70	14.93
Walk around home	0.02	0.81	12.48
<i>H</i> total: .52, <i>H</i> ^T :.44, <i>H</i> _a ^T :2.7%			
Rho = .73			n=391

Table 4.3c Mokken Scaling Procedure, Applied to 'Diabetes Only' Group

Item	Mean	ItemH	Z
Walk ½ mile	0.13	0.71	12.27
Heavy Housework	0.11	0.63	11.39
Climb 10 steps	0.09	0.75	13.19
Rising from a chair	0.03	0.88	11.71
Walk around home	0.02	0.91	10.25
<i>H</i> total: .74, <i>H</i> ^T :.48, <i>H</i> _a ^T : 0%			
Rho = .84			n=164

Table 4.3d Mokken Scaling Procedure Applied to 'Arthritis & CHD' Group

	Mean	ItemH	Z
Item			
Heavy housework	0.32	0.47	7.02
Walk 1/2 mile	0.30	0.56	8.69
Climb 10 steps	0.16	0.66	9.35
Rising from a chair	0.06	0.73	8.20
Walk around home	0.04	0.79	7.87
H total: .60, H^T :.51, H_a^T :7.2%			
Rho = .74			n=154

Table 4.3e Mokken Scaling Procedure Applied to 'Arthritis & Diabetes' Group

	Mean	ItemH	Z
Item			
Walk ½ mile	0.33	0.64	10.69
Heavy Housework	0.32	0.54	9.16
Climb 10 steps	0.18	0.64	10.02
Rising from a chair	0.08	0.74	9.04
Walk around home	0.02	0.84	6.51
H total: .64, H^T :.56, H_a^T : 3.8%			
Rho = .76			n=179

Table 4.3f Mokken Scaling Procedure Applied to 'Arthritis & COPD' Group

	Mean	ItemH	Z
Item			
Heavy housework	0.38	0.38	5.94
Walk 1/2 mile	0.28	0.55	9.75
Climb 10 steps	0.16	0.66	11.41
Rising from a chair	0.09	0.78	11.94
Walk around home	0.04	0.78	9.24
H total:.59, H^T :.45, H_a^T :10.5%			
Rho = .71			n=210

5

Markers of Preclinical Disability in Older Adults

5.1 Introduction

5.1.1 Mobility reserve

Mobility difficulties usually signal the onset of a pathway of progressive disablement (Fried et al., 1994). It is generally accepted that mobility in older adults is the first domain to decline in a three factor hierarchy, followed by instrumental and then basic activities of daily living (Fried et al., 1994; Barberger-Gateau et al., 2000). However, some item overlap has been shown with regard to instrumental activities of daily living (IADL) items always being more difficult than basic activities of daily living (ADL) items (Kelsey et al., 1998). Typical mobility tasks may include climbing stairs or walking outdoors (e.g. 800 m), while common instrumental activities might include shopping and food preparation, and finally, frequently assessed basic activities of daily living include dressing, bathing, and toileting. The prevalence of mobility limitation increases sharply with age, but as Gardner et al. (2006) observed, even in middle age (50 to 64 years) almost one-fifth of the sample reported some degree of difficulty. Wang et al. (2006) observed self-reported disability rates for community-dwelling men and women (mean age 72) to be 31% in the mobility domain, 17% in the IADL domain, and 2% in the ADL domain. Lastly, a Canadian census-based study of those 65 years and older (n = 38,518) found that 40% of respondents reported at least one disability. Perhaps more noteworthy from this census-based study, is the observation that mobility and agility disabilities accounted for 80% of all disabilities (Raina et al., 1998).

Reported mobility disability is associated with an increased risk of disability in basic Activities of Daily Living, institutionalization (Khokhar et al., 2001), and death (Melzer, Gardener, & Guralnik, 2005). For instance, in a cross sectional study, Cress et al (1995) found that older adults who need help or are unable to independently

perform activities of the gross mobility domain are at twice the risk of developing disability in ADLs. Approximately half of mobility disability develops gradually, rather than abruptly, suggesting the potential to identify individuals early in mobility decline (Guralnik et al., 2001). To prevent or reverse mobility disability, the effective targeting of interventions is essential.

5.1.2 The case for prehabilitation

Although loss of independence in everyday tasks is common in old age and rare in middle age, life-course views of disability and successful aging are increasingly recognizing that the antecedents of late-life disability may occur earlier in life (Covinsky et al., 2008). By focusing on clinical or manifest levels of ADL difficulty, the opportunities for successful intervention are minimised, because attention is targeted on recovery rather than prevention (Institute of Medicine, 2003). Several studies suggest that the prevalence of moderate disability has declined in older adults, but conflicting results exist about the trend for more severe stages (Freedman, Martin, & Schoeni, 2002), and the overall prevalence remains very high: 20% of older US adults have chronic disabilities. Furthermore, of people 65 years and older living in the community, approximately one third can be considered highly vulnerable because of advanced age, compromised functional status, frailty, and disease. This vulnerable group has been shown to benefit from clinical attention to a range of health issues, such as prevention of disease and disability and the complications of comorbid diseases, frailty, geriatric conditions (e.g., delirium, falls, or incontinence), and functional decline (Fried, 2003). Fries (2006) asks, is it possible to intervene in elderly populations, improve risk factor profiles, and observe improved health and reduced medical care costs? The answer is most likely yes; fortunately, older age is not uniformly associated with decline in performance (Seeman et al., 1994), indicating the potential for effective interventions to promote successful aging. Large randomized, controlled trials of health promotion programs in the elderly or retirees (Fries et al., 1993; Fries et al., 1994), in particular those using complex “tailored print interventions” very specific to the participant, have documented health improvement. Fries, Carey, and McShane (1997) conducted a mail-delivered arthritis self-management program. The intervention consisted of

health assessment questionnaires at three month intervals, with computer processed recommendation letters and reports individualized to age, diagnosis, education level, disability, pain, and medication. After six months of the intervention, Fries et al. obtained approximated risk reductions of 10% per year of intervention, and improved self-reported health, decreased disability and pain. Programs designed to prevent functional decline in older people show that participants with relatively good functional status or moderate frailty are those who benefit the most from these programs (Gill et al., 2002), suggesting that interventions targeted early in the process of functional decline are potentially very effective (Carriere et al., 2005).

5.1.3 Muscle strength as a predictor

Muscle strength has been reported to reach peak values between 25 and 35 yr of age, is maintained or is slightly lower between 40 and 49 yr of age, and then is ~12-14%/decade less after 50 yr of age (Lynch et al., 1999). With aging, muscle mass is lost due to motor neuron death (Metter et al., 1997) and muscle cell shrinking due to inactivity (Grimby & Saltin, 1983). Also, hormonal changes, particularly decreases in testosterone and growth hormone levels, may be associated with muscle mass decrease (Lamberts et al., 1997). However, great interindividual differences are evident in strength decline with increasing age. For example, Kallman et al. (1990) found that over an average 9-year follow-up period, 15% of the subjects ages 60 and over did not show any strength decline.

Owing largely to muscle weakness an individual may become unable to perform a regular domestic physical routine, such as rising from a chair or climbing stairs (Rantanen et al, 2002). The ability to perform normal daily household, work-related, and recreational activities is determined in part by the force-generating capacity of skeletal muscles. Muscle strength may display decremented changes with age such that a particular activity may become increasingly harder (lifting a bag of groceries), or strength may reach a threshold such that an activity can no longer be performed, e.g., standing up from a chair without assistance (Buchner & de Lateur, 1992). Young (1999) has pointed out that reductions in strength, power and oxygen consumption brought about by increasing age may mean that the performance of

some everyday tasks requires maximal effort. According to Buchner et al. (2001) strength may be curvilinearly associated with functional status so that the most evident association would be found among the very frail elderly. Among elderly persons the reserve in performance capacity may be so slight, that even a small additional decline in strength may render some everyday activities impossible.

Hand-grip strength was found to increase up until the thirties and to start to decrease with accelerated speed after the forties (Kallman et al., 1990). The Use of grip strength as a feasible marker to describe overall strength changes is supported by its significant correlation with other strength measures: $r = .638$ for elbow flexion, $r = .524$ for knee extension, $r = .515$ for trunk extension, and $r = .437$ for trunk flexion (Rantanen, Era, Kauppinen, & Heikkinen, 1994). The rates of decline in isokinetic strength averaged 14% per decade for knee extensors in men and women. Women demonstrated slower rates of decline in elbow extensors and flexors (2% per decade) than men (12% per decade). The correlation to trunk measures is relevant because trunk endurance and strength are associated with mobility performance, and have been the target of rehabilitative care (Suri, Kiely, Leveille, Frontera, Bean, 2009). More importantly, lower extremity muscle power was no better than knee-extension torque or handgrip in the early identification of poor mobility, defined either as walking speed <0.8 m/s or inability to walk at least 1 km without difficulty and without developing symptoms (Lauretani et al., 2003). Similarly, Visser, Deeg, Lips, Harris, and Bouter (2000) found that leg muscle mass was positively associated with lower-extremity performance (walk time and chair stand) in men (regression coefficient .178, $P = .035$) and approached significance in women (.202, $P = .052$). Grip strength was positively associated with lower-extremity performance in men and women. After additional adjustments for behavioural, physiological, and psychological variables, the association with leg muscle mass and lower-extremity performance disappeared, whereas grip strength remained independently associated with lower-extremity in men (.079, $P = .0001$), with a tendency in women (.046, $P = .11$). Finally, Taekema, Gussekloo, Maier, Westendorp, and de Craen (2010) found that lower handgrip strength predicted an accelerated decline in ADL disability and cognition, and thus contributes to increasing dependency in old age. Lower handgrip

strength predicted an accelerated decline in ADL disability in the functional health domain, with 0.02 points increase in the Groningen ADL Scale (possible scores range from 9 to 36 points, best to worst) per kilogramme loss of handgrip strength, $P < .01$). The range of handgrip strength was 10 to 54 kg for men and 1 to 32 for women.

5.1.4 Chronic conditions

Research in older adult populations has also demonstrated that chronic conditions are highly prevalent and, in fact, reflect the most important determinants of disability (Valderrama-Gama et al., 2002). The impact of individual chronic conditions on disability in the elderly has long been studied. Individually, musculoskeletal diseases (including arthritis), cognitive deficits, stroke, fractures, coronary heart disease and visual problems are strongly related to various functional disabilities in the elderly. However, multi-morbidity frequently occurs in older adults (Fried et al., 1999), and an increase in the number of diseases has been shown to be associated with an increase in the risk of disability for activities of daily living (Griffith et al., 2010).

Adverse health outcomes among older people are due to complex interactions of many factors (Fried et al., 2004). One component is the presence of comorbid conditions. Two thirds of Medicare beneficiaries have 2 or more chronic conditions, and rates of comorbidity increase with age. Comorbid conditions are more difficult to treat, adding complexity in terms of: competing risks; potentially incompatible therapies; burden or costs of therapies that the patient cannot tolerate; and synergistic likelihood of adverse outcomes, including disability and death (Fried et al., 2003). In relation to muscle strength mentioned above, diseases may cause decrease in strength through inactivity, or they may have a direct effect on muscle (Skelton et al., 1997). For example, in stroke, the injury in the central nervous system will affect the descending neural pathways and result in poor motor unit activation (Lindmark & Hamrin, 1995).

5.1.5 Cognition as a predictor

The Digit Symbol Substitution Test (DSST) is a subtest of the Wechsler Adult Intelligence Scales-Revised (Wechsler, 1981). DSST is a performance subtest (nonverbally mediated responses) which is commonly thought to assess motor and cognitive processing speed (Kreiner & Ryan, 2008). Digit Symbol has been frequently cited as a test that is sensitive to normal age-related cognitive decline (Birren & Morrison, 1961; Salthouse, 1992; Joy et al., 2000; Hoyer et al., 2004). Figure 5.1 on page 125 depicts a sample of the DSST test. In the DSST subtest, a set of digits and corresponding abstract symbols are presented, and the participant must match a subsequent string of digits with corresponding symbols as quickly and accurately as possible. Stephens (2006) describes the paper-based WAIS-III Digit Symbol Coding Test as: a key area that consists of a 2×9 matrix. Its first row contains the digits 1–9 numerically ordered, and its second row pairs each of these digits with a unique symbol, e.g. ‘1’ is paired with ‘–’ and 8 is paired with ‘x’. Below is a response area that consists of seven 2×20 matrices. The first row of each contains the digits 1–9 in pseudo random order, with repetitions, and the second row contains empty spaces. Participants are given 120 s to enter as many symbols as possible in the spaces below the digits according to the key area pairings. The score is the number of correctly entered symbols.

In chapter one of this thesis I noted the relationship between IQ and health status. Deary and Der (2005) demonstrated that simple and choice reaction time measures explained the association between the frequently replicated IQ and mortality relationship. For a review of the association between IQ and mortality risk see Batty et al. (2007). The Deary and Der findings were interpreted as evidence that processing speed substantially accounted for the IQ-mortality association. Hall et al. (1999) suggested, in assessing DSST abilities, a similar interpretation in their results, to the extent that DSST is part of the “processing speed” factor of the WAIS tests. Thus, the DSST relationship to mortality is relevant to this thesis in that mortality shows strong and consistent relationship with disability. In fact, evidence also exists for a direct relationship between DSST and disability. Williamson et al. (2009)

observed that DSST significantly correlated with a physical performance battery ($r = .38$, $p = .0002$), in chair stand score ($r = .26$, $p = .012$), in balance score ($r = .21$, $p = .046$). Rapp et al. (2005) found that DSST to be associated with functional disability, beyond the effects of age, gender, education, residential status, overall cognitive status, and memory.

It is clear from the literature that DSST is primarily associated with processing speed. For instance, a meta-analysis (Joy & Fein, 2001) specifically examined the relationship between the Digit Symbol and Symbol Copy (Wechsler, 1997). Symbol Copy follows the same format as Digit Symbol, except for one key component: Symbol Copy is stripped of the coding element that appears in Digit symbol. The meta-analysis revealed a mean correlation between Symbol Copy and Digit Symbol of $r = .74$. Additionally, Digit Symbol and Symbol Copy share a strong negative relationship with age, $-.68$ and $-.58$ respectively (Joy & Fein, 2001). Thus, it would appear that speed (Symbol Copy) is the strongest predictor of Digit Symbol Score, accounting for approximately 50% of the variance in Digit symbol (Joy et al., 2004). To further clarify the speeded component of Digit Symbol, Stephens (2006) used eye-movement data and video analysis to partial out the contribution of cognitive speed vs. motor speed. Stephens compared performance for two age groups (men age 20 years vs. mean age 59 years) and indicated that there existed a reliable negative correlation between key inspection latency (sum associated with the direction of gaze within the confines of 2 x 9 key matrix) and Digit Symbol score for all subjects, which suggests that cognitive speed plays a role in Digit Symbol regardless of age. Furthermore, in contrast to the younger group, the older group exhibited a reliable and marked negative relationship between writing time (sum associated with the subject's pencil being in contact with the paper and producing a mark that contributed towards production of a symbol) and overall Digit Symbol performance, thus supporting a significant role of slowed motor speed in the age/Digit Symbol relationship.

It has been ascertained that age differences account for 40-50% of the variance in Digit Symbol and only 30-40% of the variance in Symbol Copy (statistically significant difference). This implies that some factors other than speed are involved in the age-related decline in Digit Symbol scores (Joy et al., 2004). Perhaps less apparent is digit symbols ability to assess higher-order executive processes such as controlled response implementation (Wechsler, 1981), suspension of pre-potent responses (i.e., inhibition), working memory resources (maintenance), and sustained attention (Lesso-v-Schlaggar et al., 2007). Of these abilities, memory has received the most attention; significant differences in incidental paired learning on Digit Symbol have been observed in Shuttleworth-Jordan & Bode (1995), and to a lesser degree in Joy et al. (2000).

5.1.6 Association between cognition and muscle strength

There does appear to be literature supporting the relationship between cognition and subsequent muscle weakness. In a 7-year prospective cohort study (1993–2001) of 2,381 noninstitutionalized Mexican-American men and women aged 65 and older, Raji et al. (2005) observed a longitudinal association between poor cognition and greater muscle decline. MMSE score was dichotomized as less than 21 for poor cognition and 21 or greater for good cognition. Main outcomes measures were mean and slope of handgrip muscle strength over the 7-year period and incident disability, defined as new onset of any ADL limitation at the 2, 5, or 7-year follow-up interview periods. Subjects with MMSE < 21 had a significantly greater decline in handgrip strength over 7 years than those with good cognition, independent of age, sex, education, time dependent variables of depression, BMI, and medical conditions. In general estimation equation models, having poor cognition was associated with greater risk of 7-year incident ADL disability (odds ratio 2.01, 95% confidence interval (CI) 1.60–2.52). The magnitude of the association decreased to 1.66 (95% CI 1.31–2.10) when adjustment was made for handgrip strength. While these results are promising for establishing pre-clinical disability and early interventions, the results of the Raji et al. (2005) study have some limitations. It should be noted that the use of a MSSE cut scores as low as 21 means that many subjects develop ADL disability before 7 years—perhaps the 2-year mark, or even baseline. In fact, nearly

15% of the Raji et al. sample already exhibited ADL disability at baseline. The % of subjects reporting disability at the 2 and 5-year follow-up was unreported.

5.1.7 Incorporating performance and self-report measures

As early as the 1990s researchers were quite interested in clarifying the relationship between self-report and performance measures of functional status (Cress et al., 1995; Reuben et al., 1995; Kempen et al., 1996; Myers et al., 1993). These studies found only moderate associations between performance and self-report measures. However, a recent systematic review found that if the self-reported tasks were very similar to the performance measures, the correlations varied between 0.60 and 0.86. In a recent systematic review, the relationship between self-report and performance measures was examined (Coman & Richards, 2006), the authors identified 17 studies that met the inclusion criteria. The review suggests that, although comparable, self-report and performance measures usually measure different aspects of functioning. Similar evidence was provided by Chou and Macfarlane (2009), which examined the association between self-reported and performance-based measures of lower extremity functioning and found a correlation of $r = .37$. They indicated that the performance and self-report measures were, at this level of association, related to different covariates. The covariates examined included sociodemographic variables, health indicators, cognitive function, depressive symptoms, and self-efficacy. Similarly, Angel et al. (2000) indicated that, in terms of the disablement process, performance-based and self-report measures assess something different. The authors assert that each type of measure conveys useful information about functioning in complex daily environments. Despite these findings, Coman and Richards (2006) noted that a comparison based on more precise classifications of what is actually being measured in self-reported disability will lead to correlations that are considerably higher. Having said all this, the potential lack of concordance (i.e., both types of measurement yield different information), is probably best taken up as complementary rather than contradictory; both types of measures are representative indicators of limitation within the broad domain of disability (Myers et al., 1993). Furthermore, using performance-based measures to differentiate among those within

a self-reported functional status category may provide more uniform classification (Reuben et al., 1994).

The disablement process contains multiple domains (e.g., impairment, functional limitation, and disability), with self-report measures often tapping into any number of these domains. Performance measures typically assess the functional limitation domain expressed in Nagi's (1965) fourth dimension, and can include tasks like bending over, or reaching overhead. Each of the broad domains (i.e., function or disability) reflects a critical step in the 'disablement process'. The self-reported disability domain proposed by Nagi represents any restriction or inability to perform socially defined roles and tasks expected of an individual within a sociocultural and physical environment. Some of the most commonly applied dimensions of disability that fall within the environments of sociocultural and physical are (1) basic activities of daily living, which often include basic mobility and personal care; (2) instrumental activities of daily living; (3) social roles, including occupation; (4) social activities, might include attending church as well as socializing with friends and relatives; (5) and leisure activities, which might include physical recreation, reading, distinct trips, and so on (Verbrugge & Jette, 1994).

Self-reported functional status and performance measures can be used to complement each other and ultimately account for a broader range of disability variance. The benefits of performance measures are quite apparent: they have excellent face validity and reproducibility, are sensitive to change, are acceptable to patients, and focus on actual and not perceived ability. This last point however, can be a detriment, as one's ability is dependent upon their current level of motivation (Hoeymans, 1997). Another less obvious disadvantage is that performance measures usually reflect maximal ability in an artificial environment, which may not be an accurate reflection of performance in daily life (Cohen & Richardson, 2006). One mechanism that is sure to underlie the disparity between two disability indicators, is adaptive functioning; performance-based measures do not reflect adaptations made in the person's everyday living situation (e.g., compensatory mechanisms like residential structure such as type of stairs, surfaces and handrails, or avoidance of difficult

situations such as a driving a car in the dark). Such adaptive responses to physical limitations may be influenced by factors such as personality and intelligence (Kempen et al., 1996). The later factor's (i.e., cognitive ability) association with functional status has a relatively long history of replication. For instance, Bosma et al. (2007) found that the contribution of intellectual abilities (Groningen intelligence test) to functional decline (IADLs) was independent of the potentially confounding influence of early life socio-economic conditions, including childhood deprivation, early life developmental factors, such as low birth weight. It was reported that this was particularly due to the small effects of these early conditions on later functional decline (Bosma et al., 2007). Although the relationship between functional status and cognition is well documented, explicating this relationship in terms of adaptive functioning has been given very little attention.

As evidence of this complementary relationship, several contemporary investigations have sought to increase the prognostic value for subsequent disability by combining self-report and performance data. Wang et al. (2007) used gait speed and self-reported mobility difficulty to "rule out" older adults who are not at risk for physical functional decline (i.e., high probability of progressing toward ADL difficulty); for discriminating between older adults endorsing mobility difficulty and those that progress toward more severe IADL difficulty, the fastest gait speed showed the greatest validity. Using a logistic regression model Reuben et al. (2004) found that individuals with self-reported difficulty in mobility and poorer performance-based measures had progressively higher 4-year mortality rates. In a logistic regression model this relationship was significantly associated ($p < .05$) with lower risk for 1- and 4-year mortality (adjusted relative risks, .86 to .91 per physical performance unit).

The current study will investigate multiple domains, all of which have varying degrees of health inequalities for older adults. Above I have identified multiple variables from across domains of functioning that have been shown to significantly impact the disablement process. From a large sample of community-dwelling older adults, I will assess the independent influences of cognitive ability, physical performance, and self-reported mobility difficulty in the development of disability. I

will compare two groups, disabled vs. nondisabled, and then assess whether the above variables exert an influence over group membership (disabled vs. nondisabled) nearly ten years prior to the endorsing self-reported disability.

5.2 Methods

5.2.1 Study population

The data used in this chapter was drawn from the Carivascular Health Study, which was described in the previous chapter (page 87). Again, the sample was limited to those subjects aged 70 to 84 who endorsed at least one chronic condition or disease. The flow diagram for sample selection is depicted in Figure 5.2 on page 126.

5.2.2 Selected conditions

I selected a subset of the 5888 baseline sample and grouped them by the most common chronic diseases/conditions and comorbidities experienced by older adults (age 65+), as outlined by Weiss et al. (2007). In Weiss et al., 95% of community-dwelling subjects fall within the following seven groups: No disease (approximately 30% of the sample), coronary heart disease (CHD), chronic lower respiratory tract disease (CLRT), arthritis, diabetes, arthritis and CHD, arthritis and CLRT, arthritis and diabetes. Five percent of the sample presented with less common conditions, such as cerebrovascular accident or the co-occurrence of three or more diseases. Thus, the sample I chose to examine were free of stroke, vision problems, low BMI (< 18.5), current smoker status, depression scores ≥ 16 (Centre for Epidemiological Studies Depression Scale), and Mini-Mental State Examination ≤ 24 . Heart problems included a history of hardening of the arteries, heart troubles or other blood diseases. CLRT respiratory problems were limited to emphysema and bronchitis. Vision problems included eye troubles that could not be relieved by glasses. Conditions were identified by each subject's participation in standardized interviews by trained interviewers and an extensive examination at the field center. Interviews included demographic characteristics, self-assessed health status, health habits, physical activity, physical function, and medications used, and self-report of physician diagnosis of myocardial infarction, angina, congestive heart failure (CHF),

hypertension, stroke, asthma, emphysema, diabetes, intermittent claudication, renal disease, arthritis, hearing impairment, visual impairment, and cancer.

5.2.3 Predictor variables

An interviewer-administered questionnaire ascertained self-report of difficulty performing 17 specific tasks of daily life, using a modified version of the Health Interview Survey Supplement on Aging questionnaire (National Center for Health Statistics, 1987). This asked for the participant's assessment of whether he or she had difficulty with bathing, dressing, eating, using the toilet, walking around their own home, and getting out of a bed or chair (i.e. activities of daily living); heavy housework, light housework, shopping for personal items, preparing their own meals, managing their money [such as paying bills], or using the telephone (i.e. instrumental activities of daily living); with lifting or carrying something as heavy as 10 lb, walking 1/2 mile, walking up 10 steps (i.e. Rosow-Breslau scale); or with reaching out with their arms, or gripping with their hands. From this questionnaire, I extracted a mobility domain for further analysis, as observed in Fried et al. (1994). A hierarchical mobility scale was established for the six chronic conditions/diseases outlined in chapter four. I combined these subjects into one group, and then applied the Mokken scaling procedure to confirm an item hierarchy for the total sample. The combined group included 1372 men and women. The reduction includes subjects with multimorbidities, e.g., an individual with CHD, arthritis, and diabetes would not be included. After accounting for missing values in the independent variables, removing subjects that reported IADL or ADL difficulty at baseline, and death at the nine-year follow-up, the final sample was reduced to $n=710$.

In addition to using low MMSE as an exclusion criterion, I assessed cognitive ability with the Digit Symbol Substitution Test (Wechsler, 1981). Participants' function was also objectively evaluated using standardized performance based measure of strength and mobility. These examinations included triplicate measures of maximal grip strength in the dominant hand using a hand-held JAMAR Hydraulic Hand Dynamometer, as well as 15 feet walk time in seconds (usual pace).

5.2.4 Outcome measure

This study examined the influence of the independent measures (outlined above) on disability at an 8-year follow-up. Disability was defined as difficulty with any basic activities of daily living (Phelan et al. 2004; Guralnik et al., 1994), which included self-reported difficulty in bathing, dressing, eating, using the toilet, walking around their own home, and getting out of a bed or chair. Subjects reporting difficulty with ADL performance at baseline were excluded from the analysis (n=22).

5.2.5 Covariates

The following variables at baseline interview were included as covariates in the statistical models: age (continuous variable), and years of education (organised into four categories: $\leq 8^{\text{th}}$ grade; 9-11; high school graduate or some college education; and college graduate or postgraduate education). Both the Digit Symbol Substitution Test (DSST) and walk time were adjusted for height.

5.2.6 Statistical Analysis

The Mokken Scaling Procedure (version 5.0 for windows; Molenaar and Sijtsma, 2000) is a nonparametric hierarchical scaling method based on the principals of Item Response Theory. Mokken scales consist of an item selection algorithm to partition a set of items into Mokken scales and several methods to check the assumptions of two nonparametric item response theory models: the monotone homogeneity model (MHM) and the double monotonicity model (Mokken & Lewis, 1982). Mokken scaling is supported by a simple measurement model, the Guttman scalogram, which assumes that each person has a location on the latent continuum, and that a correct answer (e.g., able to perform stair climbing) occurs if and only if the person is more able than the item is difficult (Sijtsma & Molenaar, 2002). In our investigation of the latent trait (i.e., disability) we are concerned with establishing that one person has more of the latent property than another.

This chapter first applies the Monotone Homogeneity Model (MHM) to our sample of older adults, which is an IRT model used for measuring persons on an ordinal scale. Monotone homogeneity means that subjects can be ordered invariantly along

the latent scale, e.g., functional ability in the present study (Junker & Sijtsma, 2001). The MHM is based on the assumptions of unidimensionality, local independence, and monotonicity. The MHM describes item responses data that were generated by a set of homogeneous items having item response functions that are monotonically related to the latent trait. Thus we are seeking to confirm that our mobility scale measures mobility status and no other latent traits, and that the subjects can be rank ordered with respect to mobility status by means of the number of correctly answered items—in the sense of stochastic ordering. Given that the MHM holds for this set of items: the older adults with X_+ (observed sum score) = 5, for example, have a mean latent trait that is at least as high as the mean latent score of the adults with X_+ = 4; this later group has a mean latent score that is at least as high as that of the adults X_+ = 3; and so on.

A convenient way to understand the Mokken output is to first consider the relationship between the scalability of items pairs, i.e., H_{ij} . Then we consider the scalability of each item (H_i), which loosely speaking, is the extent to which item i fits together with the other $k-1$ items (Sijtsma & Moloenaar, 2002). Finally, we consider the scalability of the whole test, which is expressed with the Loevinger H coefficient (Loevinger, 1948), where higher values of H imply fewer violations and thus a better hierarchy of persons along the continuum (Kempen et al., 1997). Violations are defined as the deviation of the observed data structure from the perfect scalogram structure, i.e., an easier item wrong, or a more difficult item right (Sijtsma & Moloenaar, 2002). The scalogram assumes that each item and each person has a location on the latent continuum, and that a correct answer occurs if and only if the person is more able than the item is difficult (De Jong & Molenaar, 1987). Thus, the H coefficient is used to measure the conformity of a set of items to Mokken's criteria and validates their use together as a scale for a unidimensional latent variable (Kempen et al., 2000).

MSP also calculated z -values for each item in an effort to determine the probability of obtaining the scale. Rho was employed for the determination of scale reliability. A Rho criterion greater than .7 is thought to provide evidence for a scale's reliability.

‘Crit’ values are used to inform the researcher as to whether violations in monotonicity and double monotonicity have occurred for each item. The Crit is a summary of the H -value that informs the researcher as to the size and frequency of violations (Watson et al., 2007).

After accounting for missing values in the independent variables, removing subjects that reported IADL or ADL difficulty at baseline, and death at the nine-year follow-up, the final sample was reduced to $n=710$. The remaining data analyses were conducted using Statistical Package for the Social Sciences (SPSS) for windows (version 14.0, SPSS Inc., Chicago, IL, USA). I examined the relationship between independent measures and outcome measures with binary logistic regression. I compared endorsing difficulty with self-reported mobility items, physical performance measures, and cognitive performance at baseline with the presence of disability at the nine year follow-up. Descriptive statistics were calculated to determine the mean and distribution values for individual scale items and total scores.

5. 3 Results

As observed in chapter four, of the six chronic conditions or diseases examined with the Mokken scaling procedure, ‘diabetes only’ proved to be the only group that did not meet the monotone homogeneity model. The MHM was used to investigate three different scalability coefficients to determine if a set of predefined items form a unidimensional scale: H_i for individual items, H_{ij} for item pairs, and H for a set of items as a whole. For a set of items to be accepted as a scale, it is required that all H_{ij} pairs be greater than .00 and all H_i s be greater than or equal to .30. The ‘lift and carry groceries’ item was rejected due to a negative H_i with one of the scale items (i.e., walk in home), and thus I removed this item before combining all chronic conditions for a final analysis. As a result, the final mobility scale was reduced from six to five items.

The final scale analysis included the combining of all six groups (groups reported in chapter 4) into a sample of $n = 1372$. As noted above the final number of subjects to be analyzed was ultimately reduced to $n = 710$. With regard to the reduction of subjects that, Mokken scaling is unable to compensate for missing data, so that, all subjects with any missing item information, despite retaining information for all other items, were excluded from analysis. This particular reduction in subjects was $n = 143$, which resulted in the $n = 1229$ presented in Table 5.1 (page 127). The final scale is depicted in Table 1, and appears to be consistent with previous investigations into the mobility domain; we found that difficulty in ‘heavy housework’ was endorsed with the most frequency, followed by ‘walk ½ mile, and then ‘climb stairs’ (Wang et al., 2007; Koyano et al., 1988). Table 1 further reveals that the MHM for all items pairs was positive, which implies unidimensionality. The H coefficient for the entire scale is .65 which indicates a strong scale, well beyond the threshold of .50. Thus we can be confident that our subjects can be rank ordered with respect to mobility status. The criteria for invariant item ordering were met because the percentage of negative coefficients at the level of the individual patients (H_a^T) is 9% and the coefficient for the total set of patients (H^T) is .50, with the minimum requirement being .30. Again, the larger the H^T , the greater confidence we can have in confirming invariant ordering of items across the latent trait (Molenaar & Sijtsma, 2000).

The second phase of analyses included binary logistic regression ($n=710$), which was employed to assess the relationship between self-reported disability (dependent variable, at 8 years post baseline measurement of independent variables) and multiple predictor variables. Again, disability was defined as difficulty with one or more ADL items (bathing, dressing, eating, using the toilet, walking around their own home, and getting out of a bed or chair). The following independent variables and covariates were investigated: age (continuous), education (four categories) walk time and grip strength (both adjusted for height), and digit symbol substitution test. It should be restated that subjects were free of stroke, vision problems, low BMI (< 18.5), ‘current’ smoking status, depression scores ≥ 16 (Centre for Epidemiological Studies Depression Scale), and Mini-Mental State Examination ≤ 24 . The difference between the constant (threshold) and predicted values resulted in a chi-square value

of 131.7 with 10 d.f. The chi-square value is highly significant, which tells us that the model determinants have a significant effect on the level of reported disability.

The model indicated that four predictor variables, representing three different domains, make significant contributions to developing disability 8 years post baseline. The 'physical performance' walk time variable (after adjusting for height) indicates that a decrease in one unit (standard deviation) of walk time increased the odds of being disabled 8 years later by a factor of 1.36 [95% confidence interval (CI): 1.13, 1.64, and $p = .001$]. A trend, only, was observed for grip strength, with an 18% increase [95% CI: .662, 1.02, $p = .078$], in being disabled with a one unit decrease in grip strength. Self reported difficulty in 2 or 3 of the most difficult mobility items (as defined by Mokken scaling) increased the odds of being disabled 8 years later by a factor of 3.5 [95% CI: .738, .992, and $p < .01$]. The odds of being disabled fell to 1.9 [95% CI: 1.22, 3.07, and $p < .01$] if difficulty in one self-reported mobility item was recorded. Lastly, those in the lowest quartile of Digit Symbol Substitution Test (as compared to top quartile) had a 2.2 [95% CI: 2.07, 5.18, and $p < .01$] increase in the odds of being disabled 8 years later. Figure 5.3 on page 128 presents the percentage of disabled subjects vs. healthy subjects, with each bar representing approximate quartiles of digit symbol scores. Because the mean, median, and mode of this sample fell within the WAIS-R normed data for ages 70-74, I thought it might be informative to organise digit symbol performance by the normed data. Bar number one reflects the worst digit symbol performers (scaled scores 4-9; 1-37%), and includes 19% ($n = 136$) of the current Cardiovascular Health Study (CHS) sample. Bar number two includes the normed mean (scaled scores 10-11), and includes 30% of the current CHS sample. Bar number three includes scaled scores 12-13, and bar number four presents the best performers (scaled scores 14-18; 90-100%) which includes 26% of the CHS sample. The graph indicates that 37% of subjects have an OR of 2.7 [95% CI: 1.43, 4.97, and $p = .002$] for being disabled when compared to the top performers in bar 4. Figure 5.4 (page 129), as in Figure 5.3, presents the percentage of disabled vs. healthy subjects by quartiles of digit symbol performance. However, unlike Figure 5.3, Figure 5.4 does not provide information relating to scaled scores. In Figure 5.4, bar 1 reflects the worst digit symbol performers and bar 4 indicates the best digit symbol performers. I thought

this might be useful or informative because one could use an individual's peer derived age-scaled score to speculate (despite the fact that the above analyses does not prove causality) as to the odds of that person becoming disabled.

The final set of results relate to potentially identifying a system profile that would suggest a high risk group by the existence of simultaneous impairments. When I grouped together the baseline co-impairments of self-reported difficulty with one of the three most difficult mobility items (nearly 20% of sample), digit symbol performance (bottom 20% of CHS sample), and those subjects that fall one standard deviation below the mean for walk time performance, the percentage of subjects disabled 8 years later more than doubles to 44%. And if I replace self-reported difficulty in one of three mobility items with self-reported difficulty for two or three items (8% of CHS sample), then the number of subjects endorsing ADL disability increases to 55%.

5. 4 Discussion

It is clear from the literature that performance measures might capture preclinical or subclinical states of functional compromise before an individual reports difficulty in ADL and IADLs (Wang et al., 2002). This chapter also shows, as previously reported by Saitoh (2006), as well as Fried et al. (2000), that self-reported mobility difficulty can act as a preclinical disability marker. However, unlike other studies, I did not make use of a rather specialised and uncommon scale, i.e., questions pertaining to modification in order to achieve everyday activities. Disability in basic or personal care ADLs is generally preceded by difficulty in instrumental ADLs and mobility tasks (Guralnik et al., 1995). This chapter draws attention to the utility of confirming a hierarchy of functional decline, such that, the most challenging mobility tasks can serve as early indicators of future ADL disability. I used Mokken scaling procedures to identify a brief but reliable uni-dimensional scale, which appears to reflect the mobility component of the disablement process. This later point is supported by the fact that the item with the highest H coefficient, was the 'walk around home' task. This study is also useful or, even novel, in that I was able to

show that these findings hold across diverse diseases or chronic conditions, and that these conditions reflect the most common afflictions experienced by community-dwelling older adults.

Binary logistic regression revealed that self-reported difficulty in one of the three most challenging mobility items increased the odds of being disabled 8 years later (assessed by self-reported difficulty in one or more basic self-care ADLs) by a factor of 1.9. The OR increases to 3.5 if the individual endorses difficulty with 2 or 3 of the most challenging mobility item. These findings support the Stump et al (1998) study which found, for a sample of $n = 2857$, onset of ADL difficulty was most common among those with difficulty in three or more mobility items at baseline. The utility or impact of this study is defined by the fact that all of the significant indicators of disability were quite sensitive, detecting potentially at-risk subjects as much as 8-years prior to disability.

I should also reinforce, in this chapter I was also able to show that the most common chronic conditions in adults aged 70 and over follow a similar trajectory of decline in mobility performance, despite their different clinical presentations. This suggests that the broad spectrum of disorders that are common to older adults might benefit from a relatively generic mobility-based intervention aimed at attenuating or preventing disability.

Two other domains proved to exert an influence over the development of disability—namely, physical performance and cognition. The performance walk time variable indicates that a decrease in one unit (standard deviation) of walk time increased the odds of being disabled 8 years later by a factor of 1.36, and a trend was observed for grip strength, with an 18% increase [95% CI: .662, 1.02, $p = .078$] in being disabled with a one unit decrease in grip strength. This finding was consistent with Shinkai et al. (2000), which found that gait speed performed better than grip-strength in predicting ADL disability in community-dwelling older adults.

With regard to the cognition domain, those in the lowest quartile of Digit Symbol Substitution Test (as compared to top quartile) had a 2.2 [95% CI: 2.07, 5.18, and $p=.002$] increase in the odds of being disabled 8 years later. Converging evidence (Stephens, 2006; Deary & Der, 2005) supports the notion that cognitive speed, as opposed to motor speed, accounts for the largest part of the age-related decline in processing speed. Furthermore, higher order executive functions associated with memory play a lesser but significant role in digit symbol performance (Joy et al, 2004). The results presented in Figures 5.3 and 5.4 are noteworthy in that they seem to reflect a threshold effect for cognition (as measured by digit symbol) and disability, rather than a linear relationship. This is consistent with Guimaraes (2007): when health capital (e.g., cognitive reserve) falls below a certain level it results the crossing of a disability threshold. This threshold effect has been noted in the physical domain (muscle strength) of the functional limitation stage of the disablement process (Rantanen et al., 1998).

The presence of co-impairments has been shown to significantly increase the relative risk of walking disability. For instance, Rantanen et al. (2001) reported that the relative risk for developing walking disability was more than 5 times greater in the group with poorest balance and strength (RR 5.12, 95% confidence limit 2.68, 9.80) compared with those who had the poorest balance and best strength, the RR was 3.08 (95% CI 1.33, 7.14). For those with the best balance but poorest strength, the RR was 0.97 (95% CI 0.49, 1.93).

The most intriguing findings from this chapter relate to co-occurrence of impairments. Hajjar et al. (2009) described the identification of a novel aging phenotype by the co-occurrence of low gait speed, high depressive symptoms and impaired executive function that resulted in a significant increase in high blood pressure and cardiovascular risk. When I grouped together the baseline co-impairments of self-reported difficulty with one of the three most difficult mobility items, digit symbol performance, and poor walk time performance, the percentage of subjects disabled 8 years later more than doubles to 44%. Additionally, if I replaced

self-reported difficulty in one of three mobility items with self-reported difficulty for two or three then the number of subjects endorsing ADL disability increases to 55%.

Disease marks the first step in the disablement process (Guralnik & Ferrucci, 2009), but the effect of diseases on a person's functioning could be modified by the individual's psychological and cognitive status, life-style, social supports, and other factors in the environment (Verbrugge & Jette, 1994). In other words, despite the introduction of pathology (in terms of the Nagi scheme) during the aging process, individual differences in reserves of strength or mobility may serve to attenuate the deleterious effects of pathology. Also, maintaining a, for instance, cognitive reserve may aid in adaptive functioning which then acts to delay the development of disability. As early as the 1990s, the perceived inevitability of age-related decreases in functional ability was challenged by a growing awareness of the heterogeneous nature of the aging process (Seeman et al., 1994). While there are indeed age-associated increases in the risks of disability, the actual onset of such problems is neither inevitable nor uniform (Seeman et al., 1995). This position is supported by both decline and improvements in performance across the life-span. Acknowledgement of the variability in rates of decline in functioning has brought new attention to questions regarding the identification of factors that are associated with more successful maintenance of functional abilities with aging.

In considering disability at 8 years post baseline, many of the CHS subjects lie beyond the influence of terminal decline, and thus would allow several years of interventions aimed at bolstering reserve in various domains. Terminal decline is a hypothesis used to explain why cognition undergoes a period of decline in the last years of life; terminal decline is thought to be intertwined with age-related declines in cognitive ability (Wilson et al., 2003). Although the benefits of cognitive training have yet to be proven, particularly as it relates to the generalizability of treatment effects (Sitzer et al., 2006), some researchers have observed that cognitive-type interventions can improve performance in activities of daily living. For instance, Ball et al. (2007), using a speed of processing training program, found improvements in IADLs that lasted no less than two years. And the beneficial effects of strength

training are less controversial. Additionally, high risk individuals in our sample may benefit from chronic disease self-management classes (Lorig et al., 1999; Phelan et al., 2004).

5.5 Conclusion

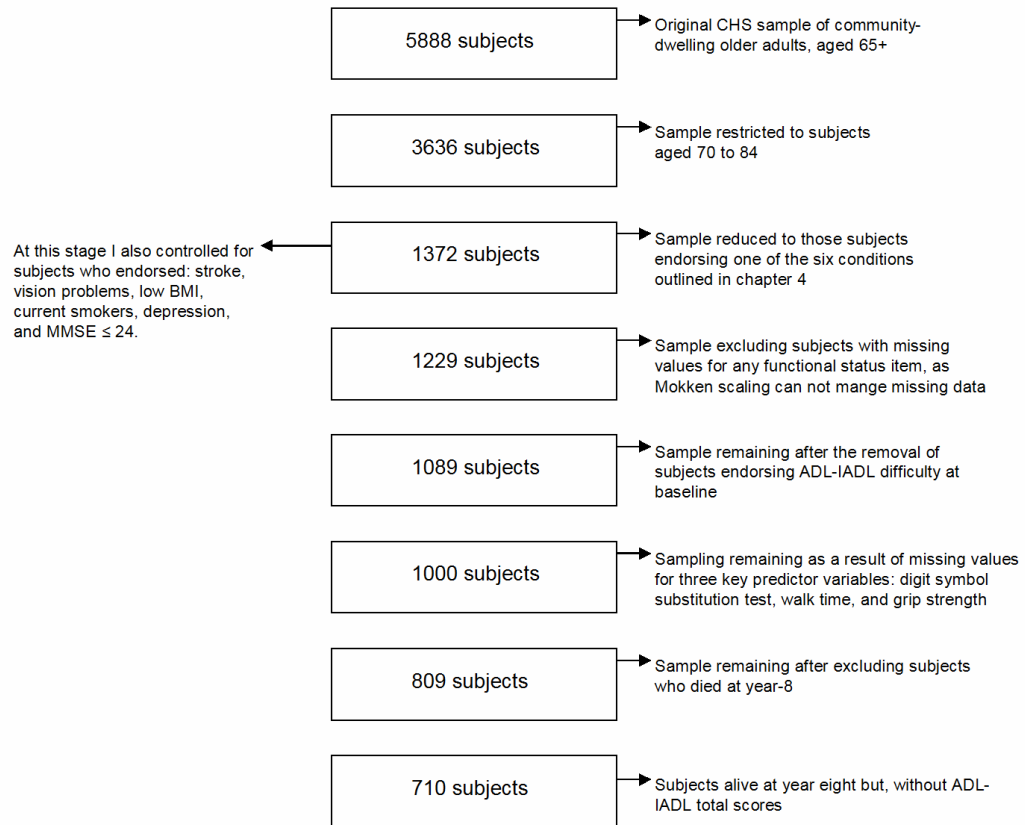
This chapter is primarily concerned with clarifying the sensitivity of methods used for the identification of those who are at high risk for disability. The preclinical disability stage is defined as: the period between the onset of impairment (i.e., loss or abnormality of anatomical, physiological or psychological structure or function) and the onset of disability, i.e., limitations in performance of actions tasks and activities expected in certain roles (Richardson et al., 2008). This chapter provides further evidence to the notion that disability manifests itself years before actual onset. I found that performance measures, cognitive ability, and difficulty in self-reported mobility all significantly increased the odds of being disabled for those free of disability at baseline. Pre-clinical disability markers have the potential to provide geriatricians with early diagnostic information pertaining to individual differences in aging, which may allow for the artificial levelling of the variability seen in healthy life expectancy.

Figure 5.1 Depiction of paper-based WAIS-III Digit Symbol Substitution Test

1 —	2 ⊥	3 ⊐	4 L	5 U	6 O	7 △	8 X	9 =	SCORE <div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto;"></div>
--------	--------	--------	--------	--------	--------	--------	--------	--------	--

SAMPLES																									
2	1	3	7	2	4	8	2	1	3	2	1	4	2	3	5	2	3	1	4	5	6	3	1	4	
1	5	4	2	7	6	3	5	7	2	8	5	4	6	3	7	2	8	1	9	5	8	4	7	3	
6	2	5	1	9	2	8	3	7	4	6	5	9	4	8	3	7	2	6	1	5	4	6	3	7	
9	2	8	1	7	9	4	6	8	5	9	7	1	8	5	2	9	4	8	6	3	7	9	8	6	

Figure 5.2 Flow diagram of subjects from the Cardiovascular Health Study



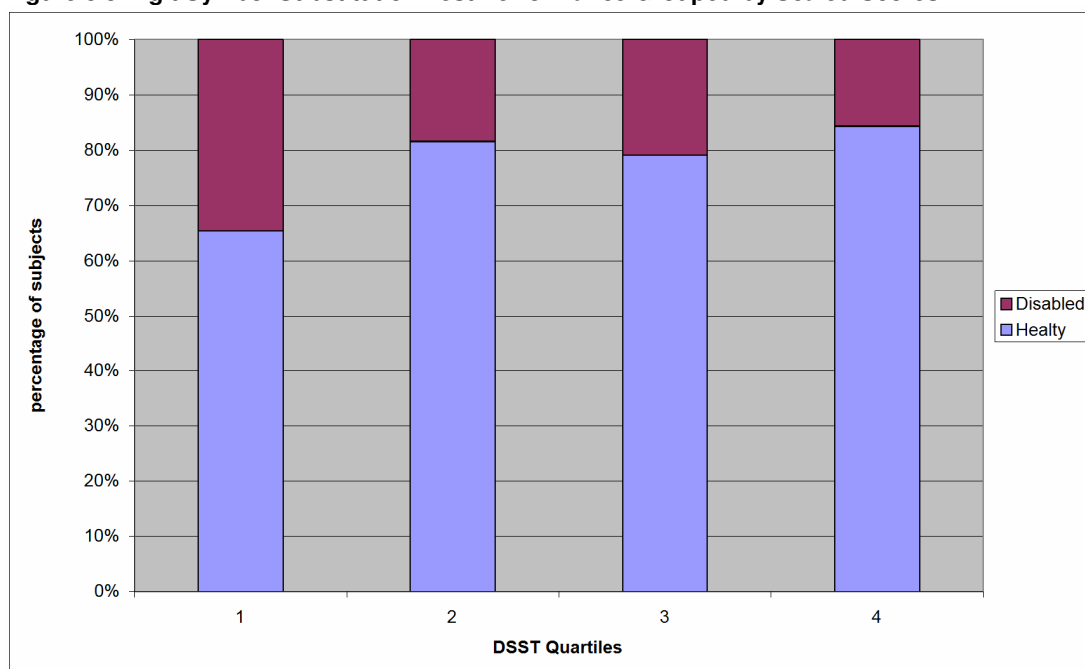
Note: BMI= body mass index. MMSE= mini mental state exam.

Table 5.1 Hierarchy for older adults endorsing chronic disease/conditions

	Mean	ItemH	Z
Item			
Heavy Housework	0.25	0.52	21.45
Walk Half Mile	0.20	0.61	26.97
Climb 10 steps	0.12	0.65	27.22
Rising from a chair	0.05	0.83	27.54
Walk around home	0.02	0.94	22.62
Scale H		0.65	
Rho = .77			n=1229

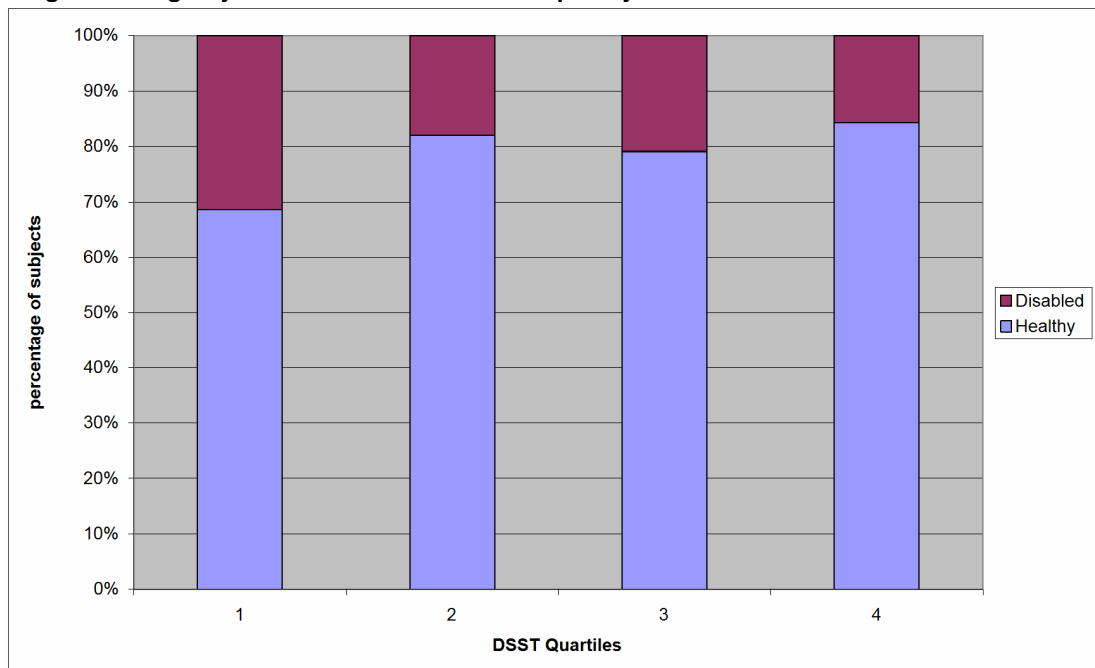
MSP = Mokken Scaling Procedure; Range for each item is zero or one (zero reflects no impairment & one indicates difficulty). Total H represents the H coefficient for the entire scale. $H^T H_a^T = HTrans$ coefficients. Rho=Mokken reliability statistic.

Figure 5.3 Digit Symbol Substitution Test Performance Grouped by Scaled Scores



Note: DSST=digit symbol substitution test. Bar 1(19% of sample) indicates worst digit symbol performers, bottom 37% of normed sample. Bar 2 (30% of sample) reflects the mean of the normed data. Bar 3 includes 24% of the sample. Bar 4 (26% of sample) represents the top 90th percentile of the normed data provided from the WAIS-R.

Figure 5.4 Digit Symbol Substitution Test Grouped by Quartiles



Note: DSST=digit symbol substitution test. Bar 1 indicates bottom 25% of DSST performance. Bar 4 indicates the best performing quartile of subjects.

6

The role of cognitive abilities and intra-individual factors in the disablement process

6.1 Introduction

In this chapter, I take a more unified approach to the determinants of disability by speaking in terms of the disablement process proposed by Verbrugge & Jette (1994). Figure 6.1 on page 146 depicts an adaptation from the Verbrugge and Jette model used in Fauth et al. (2008). In this chapter I pay particular attention to the functional limitations stage/domain and the intra-individual factors domain. This chapter also differs from Chapter five in that the variables I investigated are somewhat different. For one, the outcome measure here is a more traditional ADL-IADL scale, rather than the mobility scale presented in chapter five. Also the cognitive variable investigated here is a fluid intelligence measure, rather than the primarily speed of processing measure employed in chapter five. As I will discuss later, such an investigation is well warranted because of the frequently replicated association between fluid abilities and the aging process. Furthermore, in this chapter I formally investigate the contribution of intra-individual factors, rather than simply controlling for them as I did in chapter five. Intra-individual factors can include overt changes in lifestyle, activity level, and behaviour as a reaction to disease diagnoses, psychosocial attributes and coping mechanisms including positive affect, emotional vigour, prayer, locus of control, cognitive adaptation to one's situation, having a confidant, peer groups, etc. (Verbrugge & Jette, 1994). One additional difference between chapter five and the current chapter relates to investigating the influence of cognitive reserve, which I introduced in chapter one.

6.1.1 Cognitive Reserve

As I previously discussed the reserve construct in chapter one, I will provide only a brief introduction in this chapter. The reserve construct was proposed over a decade ago to describe the capacity of the brain to cope effectively with neuropathology so as to minimize the clinical manifestations associated with such pathology. Two components are thought to contribute to reserve capacity, cerebral (passive) and cognitive (active). The passive hypothesis stipulates that protection against the consequences of brain damage is mediated by greater intracranial volume or having more neurons and synaptic connections. Active reserve, on the other hand, is characterised by cognitive processes helping to compensate for neurodegenerative changes (e.g., through alternative neuronal pathways), usually estimated by high education, cognitively demanding occupations, or being engaged in leisure activities. Previous reports have demonstrated that the functional reorganization associated with reserve can already be detected in healthy populations. This study assessed community-dwelling older adults' levels of reserve using a measure of intelligence at age 11 years, which arguably is a passive measure of reserve, and education as well as occupation, which are traditionally classified as active reserve components (Stern, 2002).

6.1.2 Fluid Intelligence

The Raven's Progressive Matrices are a collection of standardized items and the test consists of geometric analogy problems in which a matrix of geometric figures is presented with one entry left blank. The correct missing entry must be selected from a set of answer choices. The problems in the test are presented in the form of 4x4, 3x3, or 2x2 matrix of geometric figures or patterns. Figure 6.2 (page 147) depicts a typical 3x3 matrix encountered in the *Advanced* Progressive Matrices (Raven, Raven, & Court, 1998). The advanced form of the matrices is more challenging than the standard, and is most appropriate for adults and adolescents of above-average intelligence.

6.1.3 Depression

It is important to gain a better understanding as to why impairments relating to diseases and chronic conditions lead to disability in some people but not in others. Arguably, one of the most influential factors in the disablement process relates to depressive symptoms (Stuck et al., 1996). In Chapter one, I briefly discussed the relationship between depressive symptoms and disability. Depressive symptoms in older adults play a primary role in the disablement process, with findings frequently replicated and quite often producing relatively large effect sizes. Koopmans and Lamers (2006) report that the primacy of the psychosocial domain in the disablement process is supported by the fact that functioning and well-being of persons with depressive symptoms is *reduced* to levels lower or comparable to patients with chronic medical illnesses, such as diabetes or arthritis (Koopmans & Lamers, 2001). Furthermore, when mental distress (depression or depressive complaints, anxiety) is comorbid to chronic medical illness, the prognosis of a medical condition becomes more unfavourable and treatment more complicated (Black et al., 2003; Joynt et al., 2003). Thus, depressive symptoms are comparable in severity, as well as being independent of disease pathology. That is, depressive symptoms can negatively impact functional status (with a similar intensity) as common diseases, e.g., coronary heart disease or lower respiratory tract disease. In fact, Femia et al. (2001) indicated that psychosocial variables relating to intra-individual factors are just as relevant to the disablement process as impairments and functional limitations. The importance or centrality of the intra-individual factor should not be understated, as the variables contained within are thought to speed up or slow down the time taken to become disabled (Verbrugge & Jette, 1994).

Perhaps the most noteworthy finding with regard to older adults relates to the observation that major depression is less frequent compared with younger adults (Hasin et al., 2005). This finding may, in part, be associated with improved emotional regulation in older adults. Older adults report that they focus more on self control of their emotions and rate their emotion-regulation skills as better (Lawton et al., 1992; Gross et al., 1997). When dealing with an upsetting interpersonal situation, older adults report being less likely to engage in destructive behavioural responses

such as shouting or name calling (Birditt & Fingerman, 2005). Furthermore, evidence exists to suggest that older adults are able to dissipate negative affect more effectively than younger adults (Carstensen et al., 2000).

Despite the reported psychosocial improvements in older adults, there does appear to be unique psychosocial consequences related to the aging process. For one, the risk of depression in the elderly increases with the presence of multiple illnesses. Estimates of major depression in older people living in the community range from less than 1% to about 5%, but this rises to 13.5% in those who require home healthcare and to 11.5% in elderly hospital patients (Hybels & Blazer, 2003). A systematic review of community-based studies of the prevalence of depression in later life (55+) found that: major depression is relatively rare among the elderly (weighted average prevalence 1.8%), minor depression is more common (weighted average prevalence 9.8%), and depressive syndromes deemed clinically relevant yielded an average prevalence of 13.5% (Beekman et al., 1999). The finding most relevant to this study relates to subsyndromal depression; an estimated 5 million people have subsyndromal depression, symptoms that fall short of meeting the full diagnostic criteria for a disorder (Alexopoulos, 2000), with 8 to 20% of older adults in the community and up to 37% in primary care settings suffering from depressive symptoms (U.S. Department of Health and Human Services, 1999). However, if one considers covariates associated with gender, education, physical illness and bereavement, then depressive symptoms actually decrease with frequency (Blazer, 2003). Yet, two of these variables (i.e., physical illness and bereavement) are known to increase with age, and thus act as determinants of depression rather than confounds. The increased awareness of age-related risk factors has lead to the classification of late-life depression (see Figure 6.3, page 148).

Late-life depression can be accompanied by vascular risk factors (Hickie et al., 2001), as well as neurological deficits (Krishnan, 2002). Neurological findings including white matter hyperintensities or leukoencephalopathy are reported as common among late-onset but not among early onset depression patients (Fiske et al, 2009). Using longitudinal data, Teodorczuk et al. (2007) found that severity of

baseline white matter changes predicted subsequent depressive symptoms at 1 year, even after controlling for baseline depressive scores, quality of life assessments, worsening disability, incident stroke, educational level and mini mental state exam score. Many studies have reported that older adults with late-onset depression are more likely to have concomitant cognitive deficits, especially executive cognitive functioning deficits, or are more likely to subsequently develop dementia (Schweitzer et al. 2002). The former has lead to the geriatric-specific variant, depression-executive dysfunction syndrome (Alexopoulos, 2005). Additionally, sleep disturbances, fatigue, psychomotor retardation, loss of interest in living, and hopelessness about the future may be more prevalent in late-life depression than in depression in younger adults (Christensen et al., 1999). Finally, Fiske et al. (2009) point out that biological risk factors are particularly important in old age, largely because of age-related changes that make them more common in older adults. These risks include endocrine, inflammatory or immune factors and, as previously mentioned, cardiovascular and neuroanatomical factors.

This chapter is concerned with identifying individual differences in the aging process, and how these differences influence the likelihood of being disabled. I am particularly interested in the psychosocial and cognitive components of disablement. More specifically, I will attempt to isolate the contributions made by depressive symptoms as well as fluid intelligence. This chapter will also seek to identify whether a significant protective effect of cognitive reserve can be identified. Cognitive status is a potential indicator of impairment and was discussed as such in Verbrugge and Jette's (1994) original model. However this factor, reserve in particular, has been overlooked in explicit investigations relating to the disablement model (Peek et al., 2003). I also consider the role of performance measures, which historically, at multiple stages, have been shown to be very influential in the progression toward disability. Finally, because I have data covering a time span of 8 years I will be able to provide commentary on change scores.

6.2 Methods

6.2.1 Study population

The Scottish Mental Survey 1932 (SMS 1932) was established to test the mental ability of all people born in 1921 that attended Scottish schools on June 1, 1932. The study sought to obtain information about the distribution of intelligence throughout Scotland. The psychometric test used to assess cognitive ability was one of the Moray House Tests: No. 12. In 1999 the original participants of SMS 1932 were recruited to take part in a study related to healthy aging, which later came to be identified as the Lothian Birth Cohort Study of 1921 (LBC 1921; Deary, Whiteman, Starr, Whalley, & Fox, 2004). In all, 549 subjects who fulfilled study criteria and agreed to participate in the LBC 1921 study. Each participant retook the Moray House Test that they originally sat at age 11, which now acts as a relatively rare marker of cognitive reserve. All participants who had completed the initial reassessments at mean age 79 years (old age baseline), excluding those who had withdrawn or were known to have died, were invited to participate in the follow-up measurements at mean age 87. The flow diagram (Figure 6.4) for the LBC sample selection is depicted on page 149. At mean age 87, 268 participants were invited, and 207 were tested. Reasons for not attending included withdrawal ($n = 4$ at age 87), inability or refusal to participate ($n = 42$ at age 87), having moved away ($n = 4$ at age 87), exclusion due to dementia or memory problems ($n = 3$ at age 87), and death ($n = 8$ at age 87) (Gow et al., 2008; Starr et al., 2010).

At baseline, all participants lived independently in the community and most were in good general health. Exclusion criteria included a history of terminal systemic disease, psychiatric disturbances, and neurological disorders (such as epilepsy, dementias and multiple sclerosis). Participants who have recovered from stroke were *not* excluded. Written informed consent has been obtained for all participants after being made aware of the study procedures as well as potential risks and benefits. The Multi-Centre Research Ethics Committee for Scotland and the Lothian Research Ethics Committee approved the study. Descriptives of the sample at age 79 and follow-up period are presented below in Table 6.1 on page 150.

6.2.2 Measures

Three variables were used as proxies for cognitive reserve, which included IQ measure at age 11 (Moray House Test), number of years in full time education, and participant's level of occupation. The participants sat the Moray House Test at age 11. Children were allotted 45 minutes to complete the intelligence test, and could potentially achieve a maximum score of 76. Occupational social class was coded according to the Classifications of Occupations (General registration Office: Census, 1951), which ranged from 1 (highest) to 5 (lowest). For married women the social class of the husband was used.

At ages 79 and 87, subjects completed Raven's Progressive Matrices (Raven, Court & Raven 1977). Raven's Matrices includes 60 non-verbal items (performance was the number of correct responses), thought primarily to assess inductive reasoning ability. The LBC 1921 subjects were given the 20 minute time-limited version of the test.

Depressive symptoms were assessed with the Hospital Anxiety and Depression (HADS) rating scale (Zigmond & Snaith, 1983). The scale investigates 'depression' with the following questions: "Do you take as much interest in things as you used to? Do you laugh as readily? Do you feel cheerful? Do you feel as if you are slowed down? Have you lost interest in your appearance? Do you feel optimistic about the future?, and can you enjoy a good book, radio or TV programme?" Each question provides 4 Likert-type response options. Scores of 0-7 are considered normal, 8-10 borderline and 11 or over indicating clinical 'caseness'. The HADS was administered by an LBC research associate.

Performance tests were administered to subjects at age 79 and 87, and included 6-meter walk time (time in seconds at normal pace) as well as a measure of grip strength (measured in kilograms with the Jamar Hydraulic Hand Dynamometer) which included the best of three trials for the dominant hand.

The Townsend Disability Scale was used to quantify the final stage of the disablement process. This is an activities of daily living (ADL)-instrumental activities of daily living scale (IADL) used to assess functional capacity in the older adults. The Townsend Disability Scale (Townsend, 1962) is a self-report scale that covers nine activities (washing all over, cutting toe-nails, getting on a bus, going up and down stairs, doing heavy housework, going shopping and carrying heavy bags, preparing and cooking hot meals, reaching an overhead shelf and tying a good knot in a piece of string. See appendix, figure 1, for a sample of the Townsend scale. Each activity can be scored with a 0 (no difficulty), 1 (can do but with difficulty) or 2 (not able to do) and then scores are summed to give a total score. Figure 6.5 (page 151) depicts the primary variables in the LBC cohort examined within the disablement model.

6.2.3 Statistical analyses

Using Mokken scaling analysis (for functional status at age 87), a formal hierarchy of difficulty was established, with the analyses typically being made between subjects reporting no difficulty to those endorsing problems with the most difficult item; performance on this most difficult item was used as the reference category, with approximately 40% of subjects in the group meeting traditional definitions of disabled. To investigate the hierarchical structure of the individual Townsend items, Mokken Scaling Procedure (MSP) was employed (Sijtsma, Debets, & Molenaar, 1990). Items were first collapsed to two response options, difficulty vs. no difficulty. The scale H coefficient (.57) indicated that subjects could be accurately ordered on the latent trait by means of their sum scores. The reliability of the scale was sufficient, $Rho = .75$. Furthermore the scale met the criterion for invariant item ordering, HT (a) value of 9.6 and, and an HT for the entire group = .48. The order of item hierarchy was bathing (easiest item), reaching an overhead shelf, going up and down stairs, going shopping and carrying heavy bags, and cutting toe-nails (most challenging item). Thus, endorsing bathing difficulty would indicate relatively severe functional status limitations.

All remaining analyses were conducted using Statistical Package for the Social Sciences (SPSS) for windows (version 14.0, SPSS Inc., Chicago, IL, USA). Analysis included linear regression, as well as multinomial logistic regression. A linear regression model fits a linear function to a set of data points, with the aim of minimizing the sum of the squared errors and gaining greater predictive power over a dependent variable. Just like linear regression, logistic regression gives each regressor a coefficient b_1 which measures the regressor's independent contribution to variations in the dependent variable. Although logistic regression finds a best fitting equation just as linear regression does, the principles on which it does so are rather different. Instead of using a least-squared deviations criterion for the best fit, it uses a maximum likelihood method, which maximises the probability of getting the observed results given the fitted regression coefficients.

6.3 Results

6.3.1 Direct effects on disability

The associations among seven stages of the disablement process at various ages are depicted in Table 6.2 (page, 151). The analyses were begun by examining how variables in three of the primary disablement stages (impairment, functional limitation, and intra-individual factors) influenced disability at age 87. In multinomial logistic regression I tested the influence of change in impairment (grip strength from aged 79 to 87), change on physical function (walk time from age 79 to 87), cognitive function (Raven's Matrices, 79 to 87), and change in intra-individual factors (HADS 79 to 87).

Change was evaluated by simultaneously entering each of the above variables at time point age 79 and time point at age 87. Disease status (dichotomised as any disease vs. no disease) gender, height, and BMI were also included in the model. Ultimately, the Raven's 79 and 87 variables were removed, as they were not significant and did not appear to exert any influence over other variables in the model. The same applied to HADS at age 87, and thus this variable was also removed. In the final model, change

in walk time had a significant impact on disability status at age 87 (measured with the *revised* Townsend scale). Revised refers to the total item count of the Townsend being altered to establish a formal scale hierarchy. The results indicate that with one standard unit increase in walk time, someone endorsing difficulty with two Townsend tasks compared to the most healthy subjects (those having difficulty with 0-1 tasks) had an (OR) of 2.53 [95% confidence interval (CI): 1.25, 5.14, and $p = <.01$]. Again, with one standard unit increase in walk time, the odds of being disabled increased to 4.62 ($p = <.01$) for those subjects endorsing difficulty with three Townsend tasks and then to OR 7.34 ($p = <.01$) for those endorsing 4 or 5 difficulties. Change in grip strength increased the odds of being disabled, but with a much smaller effect size. With one standard unit decrease in grip strength, for those endorsing difficulty with four tasks the (OR) was .83 [95% (CI): .715, .963, and $p = .01$]. *Change* in HADS did not significantly influence disability. However, HADS at age 79 was significant: for those endorsing difficulty with two tasks the OR was 1.39 [95% (CI): 1.11, 1.75, and $p = <.01$]. For those endorsing three tasks the OR was 1.40 ($p = .01$). Stated differently, when a participant endorses one additional depressive symptom (as compared to those with no difficulty or difficulty with one functional task), he or she has a 40% increase in the odds of having difficulty with two tasks, most likely ‘shopping & carrying heavy bags’. Lastly, and for four or five (most likely these subjects would meet the classification of disabled) tasks the OR was 1.56 ($p = .01$). That is, for one additional depressive symptom endorsed on the HADS the individual presents with a 56% increase in being disabled at the 8-year follow-up period.

I assessed reserve with three measures: IQ age 11, education and occupation. Education and occupation are commonly used to assess reserve capacity. Education and occupation attainment may reflect intellectual challenges experienced during life that enable a person to use brain networks more efficiently when confronted with insults related to the normal aging process (Katzman, 1993; Christensen, 2001; Gatz et al., 2001). Using multinomial logistic regression I observed that only education was associated with disability (Townsend measure at age 87) nearly 70 years later, although the effect was relatively small. With one year increase in education older adults were nearly 20% more likely to remain disability-free (when compared to the

reference group—subjects endorsing difficulty with at least 4 activities), with an odds ratio (OR) of 1.19 [95% confidence interval (CI): 1.01, 1.41, and $p = .05$]. However, the relationship was no longer significant after BMI was added to the model.

6.3.2 Indirect effects on disability

Despite the modest association of reserve with disability, I did find a stronger association at earlier stages in the disablement process. Within the Functional Limitations domain, one stage removed from disability, I found that education and IQ age 11 (but not occupation) exhibited a significant relationship with cognitive limitations. Using multiple regression I determined that IQ age 11 ($b = .25$, $p < .01$) and education accounted ($b = .60$, $p < .01$) for 27% of the variance in cognitive function at age 79, and 19% of the variance in Raven's Standard Matrices at age 87. I thought it might be informative to compare high reserve subjects (subjects in the 70th percentile of IQ and > 11 years of education) vs. all other subjects. Here I found that high reserve subjects scored seven points higher (raw score) on the Raven's (age 79) as compared to all other subjects, $p < .01$. This is particularly relevant, as I report later, Raven's 79 accounted for a significant part of the variance in physical limitation (mobility). Also, education, but not IQ age 11, accounted for nearly 5% of the variance in walk time (height adjusted) at age 79, with $b = .07$, $p = < .01$. Finally, passive reserve (IQ age 11) accounted for 20% of the variance in education ($b = 2.5$, $p = < .01$). This can be interpreted as one unit increase in IQ results in nearly 1.5 additional years of full-time education.

Using multiple regression, after adjusting for height, and grip strength as a covariate, depressive symptoms (HADS age 79) were significantly associated with walk time ($b = .09$, $p = .01$), and accounted for 3.3% of the variance. Similarly, after adjusting for height and gender, depressive symptoms (HADS age 79) were significantly associated with grip strength ($b = .083$, $p = .01$), and accounted for 3.0% of the variance.

The cognitive reserve measure was also significantly associated with the impairment stage of disablement and accounted for 3% of the variance in musculoskeletal ability. Musculoskeletal ability was assessed by grip strength, and adjusted for gender and height. The high reserve group as compared to all other subjects presented with a grip strength that was significant at ($t = 2.86$, $p < .01$). The β coefficient indicates a .18 increase in grip performance within a range of 6.5 (-2.3 to 4.1). Raven's Matrices (age 79) had the highest impact on functional limitation, which accounted for 6% of the variance in mobility performance (adjusted for height), as measured by 6 meter walk time (standardized $\beta = -.224$, $p = <.01$). Raven's Matrices was also associated with the impairment stage to a lesser degree, accounting for 3.3% of the variance in grip strength, after adjusting for height and gender (standardized $\beta = -.224$, $p = .080$). There was also a significant relationship between high reserve subjects and the physical component of the functional limitation stage, as measured with six meter walk time. High reserve subjects were nearly half a second ($b = .42$, $p = .05$) faster than other subjects on the 6 meter walk time test (age 79), where the average walk time was 4.3 seconds.

6.4. Discussion

This study used participants from the LBC 1921 cohort, which maintains various health related variables from age 79 and 87, thus enabling an investigation into changes in performance across time. Using this dataset I was able to demonstrate several direct effects on disability, from the impairment and functional limitation stages, on change from age 79 to 87. The outcome measure used to assess disability was a hierarchical scale adopted from the Townsend Functional Ability scale. Using grip strength to assess musculoskeletal integrity from the impairment stage, one standard unit *change* (decrease) in muscle strength from 79 to 87 appears to increase the odds of being disabled by 17%. Using walk time to assess functional limitations, one standard unit *change* (increase) mobility performance from age 79 to 87 appears to increase the odds of being disabled by 7 fold. After the inclusion of grip strength and walk time, change scores in psychosocial and cognitive variables failed to demonstrate a significant association with disability at age 87. However, a direct

effect was noted for HADS 79 to disability aged 87. This means that an LBC21 subject that endorses one additional depressive symptom at age 79, has a 56% increase in the odds of being disabled at age 87.

Fluid intelligence did exert an indirect effect on disability at multiple stages along the disablement process. Scores on the Raven's Progressive Matrices (age 79) were significantly associated with the impairment (grip strength) and functional limitation (walk time) stages. The Raven's Matrices accounted for 3.3% of variance in grip strength, with gender and height adjustments, as well as 6% of the variance in walk time. This association is made more relevant by the fact that mobility performance dramatically increased the potential for disability.

In longitudinal research, depression has been shown to be a prognostic factor of disability (Penninx et al., 1999; Preees et al., 2005). Similarly, over a period of six years Cronin-Stubbs et al. (2000), found that that depressive symptoms in community-dwelling subjects increased the likelihood of disability, with each additional symptom of depression (for the Katz-ADL measure: odds ratio, 1.16 per symptom; 95% confidence interval, 1.13-1.19; for the Rosow-Breslau measure: odds ratio, 1.14; 95% confidence interval, 1.11-1.16). After the inclusion of gender, BMI, grip strength, mobility performance, and disease severity I did not find that *change* in depressive symptoms directly affected progression toward disability. However, a direct relationship was observed from one's starting point (i.e., level of symptoms) and the odds of developing disability 8 years later. Higher depressive symptoms at age 79 increased the odds of having difficulty in 2 or more ADL-IADL tasks by 39 to 56%. Furthermore, as with previous studies (Femia et al., 2001; Peek et al., 2001), I also found evidence that depressive symptoms have an indirect association with disability progression. Depressive symptoms were indirectly related to disability by exerting an influence on the functional limitations stage of the disablement process. After adjusting for height, and using grip strength as a covariate, depressive symptoms (HADS at age 79) were shown to account for 3.3% of the variance in mobility performance as measured by six meter walk time. Depressive symptoms also accounted for 3.0% of the variance in grip strength, after adjusting for gender and height. The relationship between depression and performance may relate to the

fact that depressed individuals tend to become less active. In a longitudinal study (six years from baseline to follow-up), van Gool et al. (2003) observed a relationship between depression and physical activity; subjects who presented with ‘emerging depression’ were at higher risk (62% increase) of becoming sedentary in physical activity between baseline and follow-up.

This study appears to provide further support for the notion that cognitive stimulation (in terms of education) in one’s youth can serve to enhance resilience, with health benefits that remain measurable in later life. In examining the role of cognitive reserve, I was able to show that a presumed measure of passive reserve (IQ age 11) and a measure of active reserve, education, were significantly related to variables within the functional limitations stage of the disablement process. Employing multiple regression revealed that IQ age 11 (nearly 70 years later) and education approximately 60 years later accounted for 27% of the variance in fluid intelligence. When subjects were grouped by high vs. low cognitive reserve, using multinomial logistic regression, high reserve subjects scored seven points higher (raw score) on the Raven’s Progressive Matrices at age 79. Education, on the other hand, accounted for nearly 5% of the variance in mobility performance. In addition to its association with higher levels of cognitive performance in community-dwelling older adults, educational status has shown to be significantly correlated with both gray and white matter brain volumes, particularly temporoparietal lobes (Foubert-Samier et al., 2010). These authors concluded that the protective effect of education on brain aging may contribute to passive cerebral reserve, in addition to active cognitive reserve. The strong effect of education on aging bolsters our finding that passive reserve (IQ age 11) accounted for 20% of the variance in education. This can be interpreted as one standard unit increase in IQ results in nearly 1.5 additional years of full-time education. Lastly, education accounted for 3.3% of the variance in grip strength at age 79. The impetus for grouping subjects by high reserve vs. low reserve relates to neuroscience work associated with compensation and age-related bi-lateral activation. Cabeza et al. (2002) formed three groups of subjects based on their performance on a memory test battery, which included: California Verbal Learning Test, logical memory, and pair associates. They found that the young adults, during a source memory task, activated similar brain regions (right PFC). However, high-

ability older adults showed bilateral PFC activations. Cabeza et al. theorised that low-ability older adults, despite similar recruitment patterns to young adults, use the right PFC region less efficiently. It may also be worth noting that a trend was observed for gender and high reserve subjects regressed on BMI age 79 ($p = .075$), which accounted for 2.2% of variance.

As a group, the oldest-old (typically categorised as people over the age of 85) have the highest levels of physical disability compared with adults of other ages. However, some of the oldest-old retain high levels of functional status and avoid disability (Fauth et al., 2007). The failure to detect a direct effect between cognitive function and disability may have something to do with the overall vitality of the LBC cohort. Though it is difficult to draw meaningful comparisons between samples of older adults, from a functional status standpoint, it does appear that the LBC cohort is performing well. In a community population aged 65 years and older as much as 45% of the sample endorsed needing help, or difficulty to perform heavy housework activities (Comoni-Huntley & Brock, 1986). In another sample ($n = 4,137$ community-dwelling subjects aged 65 +) it was observed that 30% reported difficulty with heavy housework (Fillenbaum, 1985). In this LBC sample, all age 87, 44% endorsed difficulty with heavy housework. Furthermore, U.S. estimates of community-dwelling subjects aged 85 + have indicated that 22% are ADL disabled (U.S. National Center for Health Statistics, 1999). In our sample 17% could be categorized as disabled at age 87. In the 2002 English Longitudinal Study of Ageing (ELSA), of females 80 and older, 43% reported no ADL difficulty. For this LBC cohort 80% of woman reported no difficulty. With regard to men, the ELSA reported 61% of subjects with no ADL difficulty, and for the LBC men reported, 86% of subjects reported no ADL difficulty. Also many studies showing a direct cognitive relationship to disability tend to rely on global measures of cognition, such as the MMSE (Peek et al., 2003). For instance, in a Dutch study (Reuben et al., 1992) cognitive impairment was found to account for 24% of walking disability; for this study the authors derived this relationship based on $MMSE \leq 18$. Kim et al. (2005) showed a direct relationship between cognition and disability for subjects ($n=1204$, mean ages 72) that could be stratified into three groups based on their MMSE profile

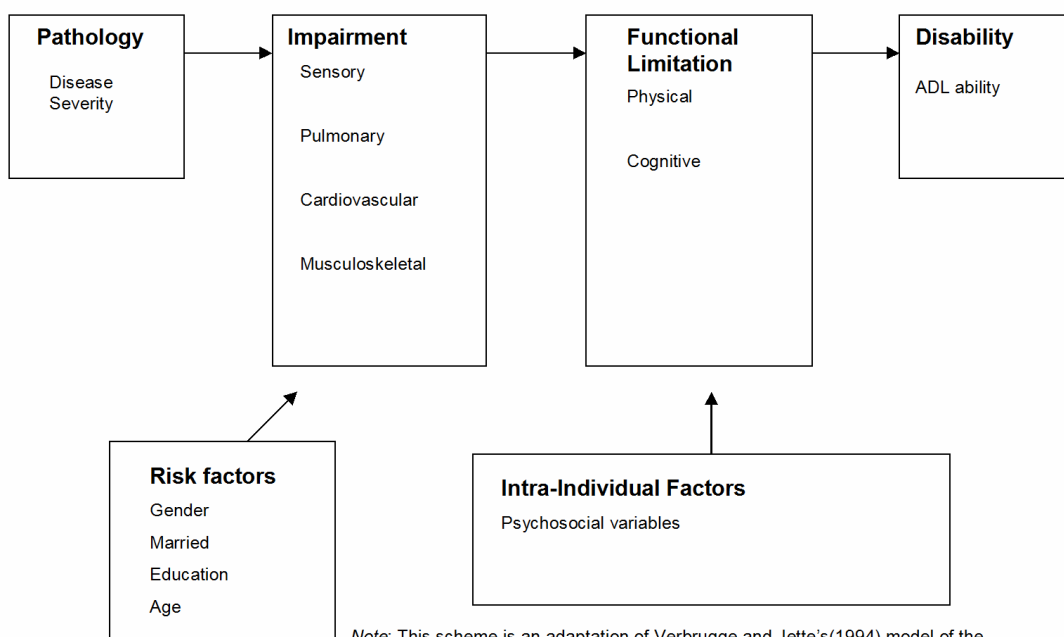
(25+, 21-24, and < 21). For the LBC sample, only 2% of subjects scored < 21, as compared to 23% in Kim et al.

6. 5 Summary

By examining functional status within the context of the disablement process, this chapter draws particular attention to the role of cognition throughout the lifespan. Cognitive function accounted for a relatively large percentage of the variance in mobility performance, and nearly seventy years prior, cognitive reserve was shown to account for a significant percentage of the physical impairment stage of disablement. In addition to cognitive ability, I was able to show that depressive symptoms act on disability at multiple stages along the pathway. Examining these variables in the light of the disablement process is made more interesting by the assertion that some factors can “speed up or slow down the pathway” (Verbrugge & Jette, 1994, p.2). For instance, change in walking performance (decline) dramatically increased the odds of being disabled, as early as eight years prior to disablement. Furthermore, fluid intelligence, indexed by Raven’s Progressive Matrices, at age 79 accounted for 6% of the variance in walking performance. Seemingly, effective interventions aimed at attenuating or preventing disability can occur at the earliest stages of development, within varying domains.

Figure 6.1

The inclusion of contextual and psychological variables



Note: This scheme is an adaptation of Verbrugge and Jette's(1994) model of the disablement process. This model builds upon the Nagi model by adding additional contextual and psychological variables. These additional variables help to explain why there is variability in the disability outcomes.

Figure 6.2 Sample item from Raven's Advanced Progressive Matrices

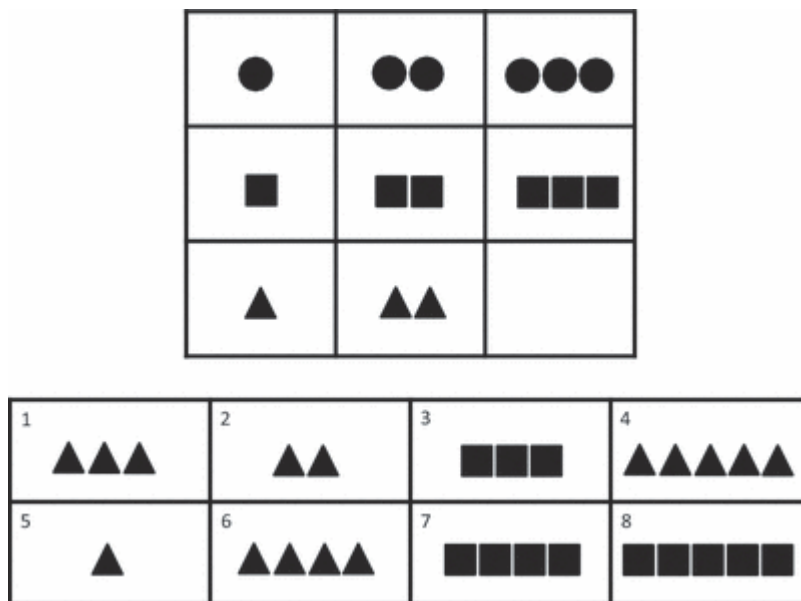
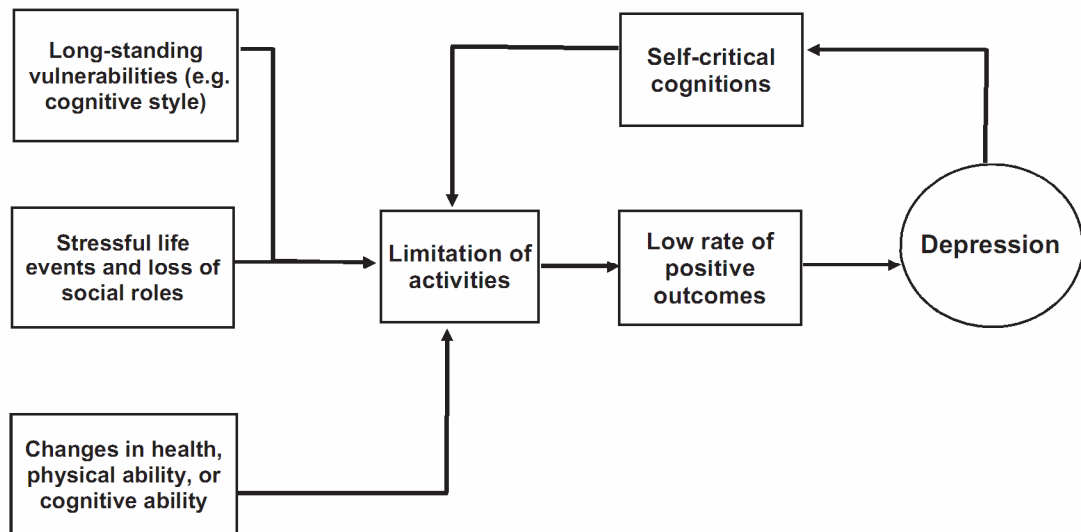


Figure 6.3 Late Life Depression Process



Note: Behavioural model (adopted from Fiske et al., 2009) reflecting the process of depression in late life.

Figure 6.4 Flow diagram for LBC study participants

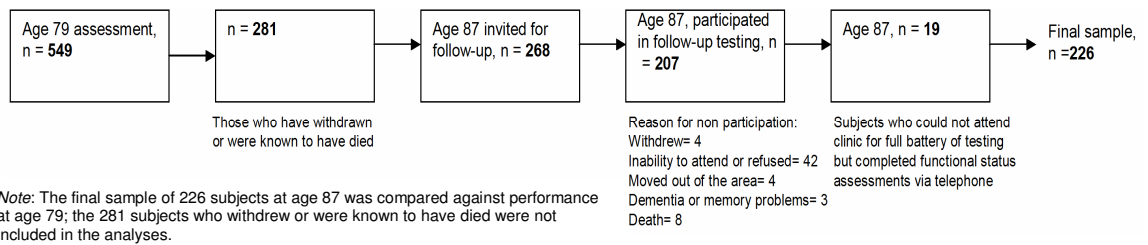


Table 6.1 Descriptive Statistics for LBC21 at Baseline and Follow-up

Variable		Baseline Died (n=322)	Baseline Survived (n=226, age 79)	Follow-up Survived (n=226, age 87)
ADL/IADL		2.5 \pm 3.0	1.9 \pm 2.3	4.1 \pm 4
Women %		40.0	46.0	54.0
MMSE score		28.0 \pm 1.78	28.4 \pm 1.5	27.8 \pm 2.2
Age 11 IQ			101.5 \pm 14.3	100.01 \pm 15.0
Raven's SPM			32.9 \pm 8.5	27.8 \pm 9.2
HADS anxiety score			4.84 \pm 3.3	4.46 \pm 3.2
HADS depression score			3.23 \pm 2.2	3.86 \pm 2.5
6 m walk time		4.97 \pm 2.2	4.37 \pm 1.4	6.8 \pm 5.1
Grip strength		25.8 \pm 8.8	27.7 \pm 9.4	21.3 \pm 8.6
FEV 1		1.83 \pm .65	1.97 \pm .58	1.77 \pm .56
Smoking status %	current		5.8	6.0
	never		47.3	38.0
	ex		46.9	47.0
Occupation code %	1		26.2	26.0
	2		37.8	38.0
	3		35.1	35.0
	5		0.9	1.0
Cardiovascular disease %	No	67.0	72.0	66.0
	Yes	17.0	15.0	25.0
	Unsure	14.0	12.0	8.0
Cerebrovascular disease %	No	92.0	90.0	87.0
	Yes	7.0	9.0	12.0
	Unsure	1.0	1.0	1.0
Diabetes %	No	93.0	97.0	95.0
	Yes	6.0	3.0	5.0
Eye disease %	No		95	48.3
	Yes		5	51.7
Systolic BP (mmHg)		163 \pm 28	164 \pm 27	157 \pm 22
Diastolic BP (mmHg)		83 \pm 15	84 \pm 13	75 \pm 12
Height (cm)		163 \pm 9	163 \pm 10	161 \pm 10
Weight (kg)		70 \pm 13	70 \pm 12	68 \pm 14
BMI		26.3 \pm 4	26.2 \pm 4	26.1 \pm 4

Note: occupation code ranges from 1 (highest) to 5 (lowest); for married woman the social class of the husband was used. BMI = body mass index. Raven's SPM=Raven's Standard Progressive Matrices. MMSE = Mini Mental Status Exam. FEV₁= forced expiratory volume.

Figure 6.5 Variables examined within the disablement model

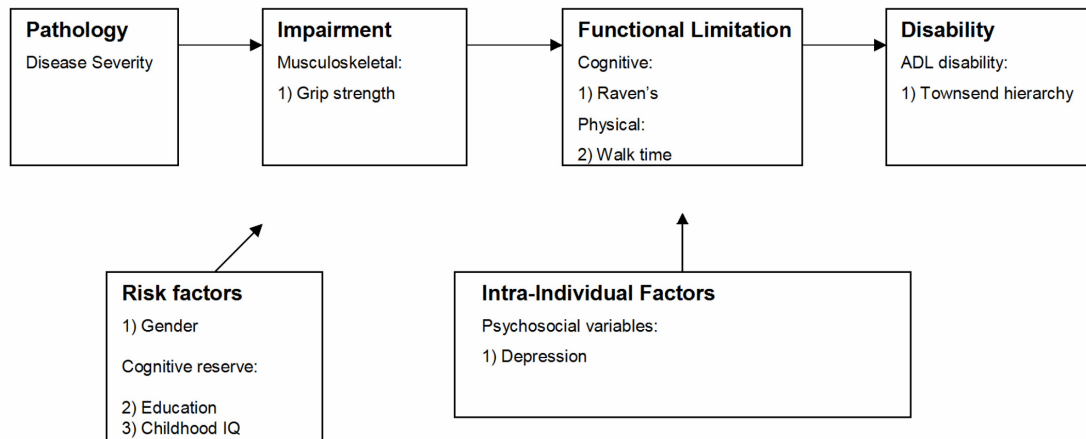


Table 6.2 Correlations between disablement components

variable		1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)	14)
Gender	(1	--													
IQ age 11	(2	.114	--												
Education	(3	-.053	.505**	--											
Occupation	(4	.056	-.401**	-.478**	--										
Disease	(5	.210**	.005	-.053	.062	--									
Raven's 79	(6	-.109	.486**	.412**	-.289**	-.023	--								
Raven's 87	(7	-.033	.378**	.335**	-.174*	-.012	.768**	--							
Grip strength 79	(8	-.839**	-.066	.126	-.074	-.227**	.192**	.122	--						
Grip strength 87	(9	-.749**	-.055	.104	-.090	-.215**	.198**	.175*	.831**	--					
Walk time79	(10	.260**	-.116	-.203	.140*	.201**	-.284**	-.252**	-.357**	-.333**	--				
Walk time 87	(11	.206**	-.122	-.134	-.004	.140	-.286**	-.277**	-.268**	.325**	.551**	--			
HADS 79	(12	-.110	.025	-.083	.038	-.20	-.30	-.004	-.017	.069	.188**	-.017	--		
HADS 87	(13	-.061	.024	-.024	.006	-.027	-.009	-.060	-.016	-.076	.135	.164*	.464**	--	
ADL-IADL	(14	.246**	-.045	-.12	.103	.237**	-.145*	-.147	-.354**	-.382	.389*	.498**	.204**	.285	--

Note: Correlations reflect Spearman's Rho. * = correlations significant at .05. ** = correlations significant at .01. ADL-IADL measure refers to Townsend Disability Scale. HADS= Hospital Anxiety and Depression Rating Scale

7

Discussion

Late-life progressive disability is thought to be a consequence of a general decline in physiological systems (Ferrucci et al., 2004). Although models of disability differ, they can be said to share a common perspectives related a continuum of functioning that proceeds from minimal, tissue or organ-level pathology to major functional limitations (one's overall ability to perform activities is limited), and them ultimately to disability or handicapped status (Seeman et al., 1994). Functional status refers to skills that are necessary for physical self-care and for performing tasks essential for independent daily living (Loewenstein and Mogosky, 1999). Most often, functional status is assessed in terms of (1) basic activities of daily living (ADL), and (2) instrumental activates of daily living (IADL). ADLs evaluate (usually by self-report) over learned or habitualized self-care activities, such as bathing or transferring from a chair or bed. IADLs on the other hand include higher order tasks such as housework or preparing meals. As compared to ADLs, IADLs are thought to require greater complexity of neuropsychological organization (Spector, Katz, Murphy, & Fulton, 1987).

Functional status is a key feature in determining one's ability to live independently within the community. As I mentioned in chapter one, ADLs are essential in making a diagnosis of dementia, relate to quality of life, and can reflect the level of economic burden required to maintain people with a disability. ADL measures have also been associated with elevated rates of mortality (Carey et al., 2004). Typically, for the elderly, ADLs decline with age. For instance, McGee et al. (1998) found that less than 5% of men and woman in the under-70 age group were severely disabled compared with over one-quarter of men and over 40% of women in the over-85 group. Yet, such findings do suggest that some elderly individuals avoid significant reductions in functional capacity, Thus, while there are age-associated increases in the risks for progression toward disability, the actual onset is neither inevitable nor

uniform (Seeman et al., 1994). The utility of identifying individuals who are ‘high risk’ for future functional decline rests on the notion that it is potentially an easier state to reverse than overt disability (Weiss et al., 2007). Health-promoting intervention programs, such as in-home visits for older adults, designed to prevent or delay dependence in activities of daily life show that participants with relatively good functional status or those in the early stages of activity restriction are those who benefit the most from these programs (Stuck et al., 2002; Guralnik et al., 2001). With this in mind, gerontological research has increasingly sought to identify community-dwelling older adults who fall within the risk zone for disability. The term preclinical disability is often used to denote this area of risk. Preclinical disability is formally defined as the period between the onset of impairment (i.e. loss or abnormality of anatomical, physiological or psychological structure or function) and the onset of disability (i.e. limitations in performance of actions, tasks and activities expected in certain roles).

This thesis sought to detect early stages of activity restriction, or the pre-clinical range along the continuum of disability. The identification of the early stages of restriction can be used to inform health-promoting intervention programmes that are meant to prevent and delay dependence in activities of daily life. To meet this challenge, I examined two cohorts composed solely of community-dwelling older adults, and for the most part, free of activity restrictions at baseline assessment. I was interested in tracking variables that have previously demonstrated associations with disability, such as cognition and physical performance measures. Ultimately, I was concerned with identifying variables that have the sensitivity to detect disability years before the overt manifestation of disability. To meet this aim I examined relatively high functioning older adults, with data that spanned nearly a decade of change in health status variables. In order to meet this aim, I also set out to examine the psychometric properties of the instruments used to define disability, namely, ADL-IADL scales.

The use of ADL-IADL measures in epidemiologic or gerontological research falls primarily into the role of outcome measure. This is partly because performance measures in older adults, and not ADLs, have proven to be quite sensitive to early

declines in functional status, and thus more suited to the identification of risk zones (i.e., pre-clinical disability or early stages of activity restriction. However, because self-reported functional status limitations and performance measures are thought to make separate contributions to disability, it may be worth enhancing the sensitivity of IADLs so that they can be effective and early predictors of disability. In this thesis, I attempted to demonstrate that item response theory could be used to improve the sensitivity and validity of ADL-IADL instruments administered to community-dwelling older adults.

Psychometrics may begin with ordinal instruments (e.g. ADLs or personality rating scales), but they often seek to establish a universally accepted continuum for rating an attribute, a continuum that behaves much like an inch-mark on a ruler (Sijtsma & Molenaar, 2002). Item response theory (IRT) attempts to replicate the precision found in the physical sciences by converting ordinal and qualitative characteristics into interval and quantitative data. A variable at the 'interval' level of measurement maintains spacing between numbers which are equal. Thus, an interval variable is one where there is an arithmetical relationship between responses (Glantz, 1992).

The most common procedure used today is the summing (Likert-scoring method) of item ratings to derive a total ADL score (Ward & Murray-Ward, 1998). The simple algebraic sum of item scores has been criticized in the past, due in part to its potential for misinference (McHorney et al., 1997); Thorndike (1904) identified problems inherent in using measurements of this type, such as the inequality of the units counted and the non-linearity of 'raw scores' (Sheehan et al., 2000). Adding ordinal scores as though they were equal interval to form a total score is considered an inappropriate mathematical manipulation (Merbitz, Morris, & Grip, 1989). Thus the results lack meaning and can result in serious misinterpretations of the data (Arnadottir & Fisher, 2008). The potential for misinference is rooted in the inequality of scale intervals (Vittengl et al., 2006; Armstrong & Sloan, 1989). Health status scores do not have ratio characteristics and thus one should not conclude that a patient with a score of 50 is twice as healthy as a patient with a score of 25 (Stucki et al., 1996). "With the priority placed on establishing interval units of measure, complementary tools for understanding the nature of scale's meaning and, more

importantly, providing a substantive context within which an individual's score on a scale may be interpreted" (Haley et al., 2004, p.52). Such tools would include the ability to establish a formal hierarchy of functional decline (i.e., invariant item ordering), as well as areas of construct under-representation.

The potential for a hierarchical scale (one with a continuum of task difficulty) to be used in the identification of older adults that fall into risk zones for disability has been, in my opinion, greatly marginalised. This is in part related to the difficulty in establishing invariant item ordering. As pointed out by Sijtsma and Meijer (2011), in the past, many researchers have reported establishing an item hierarchy through IRT procedures, but in reality failed to provide accurate evidence for such a hierarchy (e.g., Kempen & Suurmeijer 1990 or Fortinsky et al., 2003); Ligtoet et al. (2010) conveys that IIO is a strong requirement in psychometrics, and that researchers wrongly assume that fitting any IRT model implies that the items have the same ordering by difficulty for all subjects. Thus, the resulting spurious predictive power for such procedures may have served to reduce the perceived utility of investigating hierarchies of functional status. Establishing a hierarchy of disability was first proposed by Katz (1963), and in fact, ADL scales are quite reliable in the formulation of an item hierarchy. However, it has been much more difficult to establish an item hierarchy with more complex IADL items. For example, Lawton (1969), the original developer of IADL scales, reported that food preparation, laundry, and housekeeping failed to form an ordered scale.

The interpretive power of functional status scales can also be boosted in other ways. For one, these instruments need to be reformulated for higher functioning community-dwelling older adults. A large proportion of functional status scales are simply unable to obtain 'readings' for community-dwelling older adults. In chapter three I was able to show that several IRT-revised scales exhibit ceiling effects beyond 45%, when the acceptable range should be between 10 and 15%. When IRT procedures are used to convert ordinal level data, sensitivity to change or responsiveness is enhanced which allows for the detection of change where it is clinically relevant. This is particularly important when tracking progressive

disability, as compared to catastrophic disability, as the former will most likely present with smaller intervals of change. The conversion to interval level measure also serves to identify regions along the scales that are ineffective in measuring performance of community-dwelling older adults, i.e., construct under-representation.

In chapter three, I reported on a systematic review that aimed to identify functional status scales revised through item response theory, with samples that were restricted to community-dwelling older adults. The review was particularly concerned with topics related to ceiling effects (i.e., large proportion of subjects scoring at the top of the scale or are unimpaired), confirming an item hierarchy, and establishing interval level measurement properties, which could be used to identify construct-underepresentation. Several scales identified in the review were exceptional at reducing ceiling effects (e.g., Jette et al., 2002), or reducing gaps in coverage along the construct (e.g., McHorney & Cohen, 2000). However, a scale that exhibits both an exceptional reduction in ceiling effects, gaps in coverage, as well as invariant item ordering for community-dwelling older adults remains elusive. This review suggests that more work could be done to improve the validity of instruments used to assess functional status in community-dwelling older adults.

In chapter three I show that not all functional status scales are created equal, and that by using IRT methods, one can enhance the interpretive power of scales. In chapter four I was able to use the IRT property of invariant item ordering to define a common hierarchy of functional decline for the most common chronic conditions and diseases experienced by older adults. The conditions investigated included, arthritis, coronary heart disease (CHD), diabetes, co-occurrence of chronic lower respiratory tract disease and arthritis, co-occurrence of diabetes and arthritis, co-occurrence of CHD & arthritis. The results of this study demonstrate the potential for mobility reserve to be used as an early indicator (or preclinical marker) for ADL disability in community-dwelling older adults. More importantly, Mokken analyses indicated that a common hierarchy of functional decline was observed for a diverse set of conditions and diseases that are common among community-dwelling older adults.

Such an indicator could be used to identify hierarchical declines relating to severity in diverse patient populations. In chapter four I presented the benefits of establishing a formal hierarchy of functional decline by demonstrating that common medical conditions in older adults share a similar trajectory of decline. In chapter five I further supported the utility of hierarchical scale analyses by showing that ‘early’ items were strong determinates of self-reported disability at 8-year follow up. The effect size surpassed those recorded for grip strength, walk time, depressive symptoms, cognitive performance, and body mass index.

In *A Novel Aging Phenotype of slow gait, impaired, executive dysfunction, and depressive symptoms: relationship to blood pressure and cardiovascular risk* (Hajjar et al, 2009), the authors investigate the co-occurrence of impairments in three domains in a population of non-demented older adults (mean age 77.8). They were seeking to identify a system profile that would suggest the existence of a specific aging phenotype that is made distinctive by their simultaneous impairment. Using an executive measure (Trail Making Test Part B), gait speed (two 4-m walk tests), and depressive symptoms (CES-D scale) Hajjar et al. identified a subgroup (17%) that performed poorly on all of these measures. As compared to non-members of this group, the impaired group presented with elevated systolic and pulse pressure, as well as stroke. This is exactly the sort of thing I was after in chapter five, that is, identifying subjects at high risk for disability based on the co-occurrence of impairments. In addition to the Hajjar et al. study, the co-occurrence of just two variables has been shown to have predictive power with regard to identifying older adults at high risk for ‘disability’; Rantanen et al. (2001) reported that the relative risk (RR) for developing walking disability was more than 5 times greater in the group with poorest balance and strength compared with those who had the poorest balance and best strength, the RR was 3.08. For those with the best balance but poorest strength, the RR was 0.97. The results from chapter five revealed that, when grouped together, the baseline co-impairments of self-reported difficulty (one of the three most difficult mobility items), digit symbol performance (bottom 20% of CHS sample), and those subjects that fall one standard deviation below the mean for walk time performance, the percentage of subjects disabled 8 years later more than

doubles to 44%. If self-reported difficulty in one of three mobility items is replaced with self-reported difficulty for two or three items (8% of CHS sample), then the number of subjects endorsing ADL disability increases to 55%.

Chapter six of this thesis takes a more unified approach to the determinants of disability by presenting decline in terms of the disablement process proposed by Verbrugge and Jette (1994). Examining these determinants in the light of the disablement process is made more interesting by the assertion that some factors can ‘speed up or slow down the pathway’. I was able to show that both cognitive and depression variables acted to speed up the process toward disability. Depressive symptoms acted on disability directly, with lower depressive symptoms reducing the odds of being disabled almost a decade later. Cognitive function accounted for a relatively large percentage of the variance in mobility performance (6% of the variance), and nearly seventy years prior, cognitive reserve was shown to account for a significant percentage of the physical impairment stage of disablement. Performance measures associated with the impairment and functional limitation stages of the disablement process were the only variables to exhibit a *change* in performance that resulted in poorer disability status.

7.1 Limitations

My discussion of the thesis limitations will begin with the systematic review presented in chapter three. This review was flawed in that it did not have a second reviewer, which would serve to increase the objective nature of the review. Furthermore the review was relatively ineffective in establishing a common hierarchy of functional status declines in community-dwelling older adults. If a common hierarchy was established, this could be useful in clinical practice; it would allow a general practitioner to establish, with relative ease, an older adult’s position along the disability continuum, and ultimately his or her risk for future disability.

In chapter four I was able to show that a rather diverse set of chronic diseases, for the most part, follow a common hierarchy of mobility decline. However, despite meeting the minimum requirements for scale reliability, several of the scales presented with

reliability coefficients that were quite low. This, almost certainly, had to do with the number of items used in our mobility scale. Sijtsma and Emos (2011) showed that increasing their scale or questionnaire length from 5 to 20 items improved reliability (Cronbach's alpha) from .70 to .90. Additionally, I would have preferred that the sample sizes for several disease groups to have been a bit larger, as low sample sizes can adversely affect one's confidence in item calibrations. It appears that most authors will *not* endorse the use of samples below 100 subjects for IRT analyses. De Morton et al. (2008) assert that sample size depends, to some degree, on test targeting (i.e., items and persons cover a similar range of the construct under investigation), and that 100 subjects would be sufficient in such cases.

In chapter five I found that cognitive performance at baseline was associated with the odds of being disabled eight years later, with relatively large effect sizes. However, it wasn't clear how much of this relationship could be attributed to the theory of terminal decline. According to this hypothesis (Kleemeier, 1962), in the last years of life cognitive ability undergoes a period of terminal decline which substantially contributes to age-related cognitive detriment (Wilson et al., 2003). It also wasn't clear whether the odds of increased disability was associated with a deterioration over time or whether it was merely the individual's starting point that determines the predictor/disability relationship.

In the final chapter however, I was able to explicitly investigate change over time. I was able to show that change in walk time and to a lesser degree grip strength was associated with a significant increase in developing disability eight years later. However, unlike these physical performance measures, *change* in depressive symptoms did not have a significant association with disability. And yet, depressive symptoms did exhibit a relatively strong relationship with disability; the relationship was not driven by an increase over time in depressive symptoms, but rather by one's baseline level. The limitation in chapter six relates to the fact that I was unable to assess change in disability. Instead, I examined change in the predictor variables, keeping disability level constant (disability at age 87). I chose not to investigate change in disability because, despite the item hierarchy being constant from age 79

to 87, the reliability for the functional status scale at age 79 was below .70. Another limitation in chapter six relates to the disablement process itself. Unlike, Lawrence and Jette (1996) who found a causal ordering of the components or stages in the disablement process, my investigation was more ambiguous. For instance, it appeared that the early stage of disablement related to impairment (measured by grip strength) had less of an impact on disability than the later stage of functional limitation, as measured by walk time.

7.2 Conclusion

- As stated in the first sentence of the first chapter, the overarching theme of this thesis is the prevention of progressive-type disability. In the service of this theme, the thesis drew attention to or implemented modern calibration techniques in order to enhance functional status instruments used as outcome measures, but also sought to argue for the use of ADL-IADL hierarchies as predictor variables.
- The thesis was also able to provide further evidence for the identification of pre-clinical determinates of disability (or high risk zones), the presence of which can be detected as early as 8 years prior to the endorsement of overt disability. Potentially, this sort of information can be used to improve the targeting of interventions aimed at preventing or delaying dependence in activities of daily life.

References

- Ackerman, P. L., Beier, M. E., Boyle, M.O. (2002). Individual differences in working memory within a homological network of cognitive and perceptual speed abilities. *Journal of experimental psychology. General*, 131, 567-89.
- Ackerman, P. L., Beier, M.E., Boyle, M.O. (2005). Working memory and intelligence: the same or different constructs? *Psychological Bulletin*, 131, 30-60.
- Adamson, J. (2007). Disability and education - the Madonna factor? *International Journal of Epidemiology*, 36, 365-367.
- Alagumalai, S., Curtis, D., Hungi, N. (2005). *Classical test theory*. In *Applied Rasch measurement: A book of exemplars*. Edited by Alagumalai S, Curtis D. Netherlands: Springer; 1-15.
- Alexopoulos, G. S., Meyers, B. S., Young, R. C., Kalayam, B., Kakuma, T., Gabrielle, M., Sirey, J. A., & Hull, J. (2000). Executive dysfunction and long-term outcomes of geriatric depression. *Archives of general psychiatry*, 57, 285-90.
- Allen, S. M., Mor, V., Raveis, V., & Houts, P. (1993). Measurement of need for assistance with daily activities: quantifying the influence of gender roles. *Journal of Gerontology Social Science*, 48, 937-42.
- Anderson J.W. Randles, K. M., & Kendall, C. W. (2004). Carbohydrate and fiber recommendations for individuals with diabetes: a quantitative assessment and meta-analysis of the evidence. *Journal of the American College of Nutrition*, 13, 5-17.
- Andrich, D. A. (1978). Rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Angel, R.J., Ostir, G.V., Frisco, M.L. & Markides, K.S. (2000). Comparison of a self-reported and a performance-based assessment of mobility in the Hispanic established population for epidemiological studies of the elderly. *Research on Aging*, 22, 715-737.
- Arnadottir, G., & Fisher, A. G. (2008). Rasch analysis of the ADL scales of the A-One. *American Journal of Occupational Therapy*, 62, 51-60.

- Arendasy M. E., Hergovich, A., Sommer, M., & Bognar, B. (2005). Dimensionality and construct validity of a video-based, objective personality test for the assessment of willingness to take risks in road traffic. *Psychological reports*, 97, 309-20.
- Asberg, K. H. & Sonn, U. (1989). The cumulative structure of personal and instrumental ADL. A study of elderly people in a health service district. *Scandinavian Journal of Rehabilitation Medicine*, 21, 171-177.
- Avlund, K., Kreiner, S., Schultz-Larsen, K. (1993). Construct validation and the Rasch model: functional ability of healthy elderly people. *Scandinavian Journal of Social Medicine*, 21, 233-246.
- Avlund, K., Schultz-Larsen, K., & Kreiner, S. (1993). The measurement of instrumental ADL: content validity and construct validity. *Aging (Milano)*, 5, 371-83.
- Avlund, K., Kreiner, S., Schultz-Larsen, K. (1996). Functional ability scales for the elderly. *European Journal of Public Health*, 6, 35-42.
- Avlund K, Lund R, Holstein BE, Due P, Sakari-Rantala R, Heikkinen R-L. (2004). The impact of structural and functional characteristics of social relations as determinants of functional decline. *Journal of Gerontology: Social Science*, 59, 44-51.
- Babcock, R. L. (1994). Analysis of the adult age differences on the Raven's Advanced Progressive Matrices Test. *Psychology and Aging*, 9, 303-314.
- Babcock, R. L. (2002). Analysis of age differences in types of errors on the Raven's Advanced Progressive Matrices. *Intelligence*, 30, 485-503.
- Baker, F. (2001). *The Basics of Item Response Theory*. University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Ball, K., Edwards, J.D., & Ross, L.A. (2007). The impact of speed of processing training on cognitive and everyday functions. *Journals of Gerontology: Psychological Sciences and Social Sciences*, 62, 19-31.
- Barberger-Gateau, P., Rainville, C., Letenneur, L., & Dartigues, J-F. (2000). A hierarchical model of domains of disablement in the elderly: a longitudinal approach. *Disability and Rehabilitation*, 22, 308-317.
- Batty, G. D., Deary, I. J., Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Annals of Epidemiology*, 17, 278-288.

- Beal, E. & Katz, S. (2004). *Myth busters: Defying the myths of aging*. Retrieved from Benjamin Rose Institute On Aging website: <http://www.benrose.org/mythbusters/mbsidneykatz.cfm>.
- Bebbington, A.C., & Davies, B. (1995). Efficient targeting of community care: the case of the home help service. *Journal of Social Policy*, 22, 373-91.
- Beckman, A. T., Copeland, J. R., Prince, M. J. (1999). Review of community prevalence of depression in later life. *British Journal of Psychiatry*, 174, 307–311.
- Beekman, A. F. T., Deeg, D. J. H., Van Tilburg, T., Smith, J. H., Hooijer, C., & Van Tilburg, W. (1995). Major and minor depression in later life: a study of prevalence and risk factors. *Journal of Affective Disorder's*, 34, 41-49.
- Bickersteth, P., & Das, J. P. (1981). Syllogistic reasoning among school children from Canada and Sierra Leone. *International Journal of Psychology*, 16, 1–11.
- Birditt, K. S. & Fingerman, K. L. (2005). Do we get better at picking our battles? Age group differences in descriptions of behavioural reactions to interpersonal tensions, *Journal of Gerontology Series B :Psychological Sciences and Social Sciences*, 60, 121–8
- Birren, J. E., & Morrison, D.F. (1961). Analysis of the WAIS subtests in relation to age and education. *Journal of Gerontology*, 16, 363-369.
- Black, S. A., Markides, K. S. & Ray, R. A. (2003). Depression predicts increased incidence of adverse health outcomes in older Mexican Americans with type 2 diabetes. *Diabetes Care*, 26, 2822–2828.
- Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioural and Brain Sciences*, 29, 109-160.
- Bond T., Ughrin T., & Fox C. (2001) *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (book).
- Bosma, I., Vos, M. J., Heimans, J. J., Taphoon, M.J., Aaronson, N.K., Postma, T.J....Klein, L.(2007). The course of neurocognitive functioning in high-grade glioma patients. *Neuro-oncology*, 9, 53–62.
- Bossoni, S., Mazziotti, G., Gazzaruso, C., & Martinelli D. (2008). Relationship between instrumental activities of daily living and blood glucose control in elderly subjects with type 2 diabetes. *Age and Ageing*, 37, 222-5.

- Boyle, P.A., Cohen, R.A., Paul, R., Moser, D. & Gordan, N. (2002). Cognitive and motor impairments predict functional declines in patients with vascular dementia. *International Journal of Geriatric Psychiatry*, 17, 164-169.
- Boyle, P. A., Malloy, P., Salloway, S., Cahn-Weiner, D., Cohen, R. & Cummings, J. (2003). Executive dysfunction and apathy predict functional impairment in Alzheimer disease. *American Journal of Geriatric Psychiatry*, 11, 214-221.
- Boyle, P. A & Cahn-Weiner, D. (2004). Assessment and prediction of functional impairment in vascular dementia. *Expert Revised, Neurotherapeutics*, 4, 109-114.
- Brach, J. S., & VanSwearingen, J. M. (2002). Physical impairment and disability: relationship to performance of activities of daily living in community dwelling older men. *Physical Therapy*, 82, 752–761.
- Branch, L. G., Katz, S., Knipmann, K., & Papsidero, J. A. (1984). A prospective study of functional status among community elders. *American journal of Public Health*, 74, 266–268.
- Branch, J. S., VanSwearingen, J. M., Newman, A. B., & Kriska, A. M. (2002). Identifying early decline of physical function in community-dwelling older woman: Performance-based and self-report measures. *Physical Therapy*, 82, 320-328.
- Brown, M., & Holloszy, J.O. (1991). Effects of low intensity exercise program on selected physical performance characteristics of 60- to 71-years olds. *Aging*, 3, 129-139.
- Bruce, B., & Fries, J. F. (2003). *The Stanford health assessment questionnaire: a review of its history, issues, progress, and documentation*. *Journal of Rheumatology*, 30, 167-78.
- Bruce, D.G., Davis, W.A., & Davis, T.M.E. (2005). Longitudinal predictors of reduced mobility and physical disability in patients with type 2 diabetes: the Fremantle Diabetes study. *Diabetes Care*, 28, 2441-2447.
- Bruce, O., Markovic, D., & Fonarow, G. C. (2011). Disability and Incident Coronary Heart Disease in Older Community-Dwelling Adults: The Three-City Study Recent US Patterns and Predictors of Prevalent Diabetes among Acute Myocardial Infarction Patients. *Cardiology Research and Practice*, Article ID 145615, 8 pages doi:10.4061/2011/145615.
- Buchner, D. M., & de Lateur, B. J. (1991). The importance of skeletal muscle strength to physical function in older adults. *Behaviour Medicine Annals*, 13, 91–98.

- Cabeza, R., Anderson, N.D., Locantore, J.K., & McIntosh, A.R. (2002). Ageing Gracefully: compensatory brain activity in high performing older adults. *Neuroimage*, 17, 1394-1402.
- Cabrero-Garcia, J. & Lopez-Pina, J. A. (2008). Aggregated measures of functional disability in a nationally representative sample of disabled people: analysis of dimensionality according to gender and severity of disability. *Quality of Life Research*, 17, 425-436.
- Carey, E. C., Walter, L. C., Lindquist, K., Covinsky, K. E. (2004). Developmental and validation of functional morbidity index to predict mortality in community-dwelling elders. *Journal of General Internal medicine*, 19,1497-1525.
- Carpenter, P., Just, M., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97,404–431.
- Carrière I, Colvez A, Favier F, Jeandel C, & Blain H; for the EPIDOS study group (2005). Hierarchical components of physical frailty predicted incidence of dependency in a cohort of elderly women. *Journal of Clinical Epidemiology*, 58, 1180-1187.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge, England: Cambridge University Press.
- Carroll, J. B. (2003). *The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors*. In Helmuth Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. Elsevier Science/Pergamon Press.
- Carstensen L. L. Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life. *Spanish Journal of Personality and Social Psychology*, 79, 644–655.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- Cattell, R. B. (1968). Fluid and crystallized intelligence. *Psychology Today*, 3, 56–62.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.

- Chiu, Y., Fritz, S. L., Light, K. E., Velozo, C. A. (2006). Use of item response analysis to investigate measurement properties and clinical validity of data for the dynamic gait index. *Journal of Physical Therapy*, 86, 778-787.
- Chou, K. L., & Macfarlane, D.J. (2009). Relationships between performance-based and self-reported measures of lower extremity function in Chinese older women. *Topics in Geriatric Rehabilitation*, 25, 181-190.
- Christensen, H., Mackinnon, A., Jorm, A. F., Henderson, A. S., Scott, L. R., & Korten, A. E. (1994). Age differences in interindividual variation in cognition in community-dwelling elderly. *Psychological Aging*, 9, 381-390.
- Christensen, H. (2001). What cognitive changes can be expected with normal aging? *Australian and New Zealand Journal of Psychiatry*, 35, 768-775.
- Classical test theory vs. item response theory: an evaluation of the theory theory in the Swedish Driving License Test [www.umu.se/edmeas/publikationer/pdf].
- Colom, F, Vieta, E. (2004). A perspective on the use of psychoeducation, cognitive-behavioral therapy and interpersonal therapy for bipolar patients. *Bipolar Disorder*, 6, 480-6.
- Coman, L., & Richardson, J. (2006). Relationship between self-report and performance measures of function: a systematic review. *Canada Journal on Aging*, 25, 253-270.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed and general fluid intelligence. *Intelligence*, 30, 163–183.
- Corti, B., Donovan, R.J., A & Holman, C.,D., J. (1996). Factors influencing the use of physical activity facilitates : results from qualitative research and health promotion. *Journal of Australia*, 7,144-142.
- Coultas, D. B., Gong, H., Jr., Grad, R., Handler, A., McCurdy, S. A., Player, R., ... Westley, M. (1994). Respiratory diseases in minorities of the United States. *American Journal of Respiratory & Critical Care Medicine*, 149, 93-S131.
- Covinsky, K.E., Lindquist. K.,Dunlop, D.D., Gill,T.M., & Yelin, E. (2008) Effect of arthritis in middle age on Older-age functioning, *Journal of the American geriatrics society*, 56, 23-28

- Cress, M. E., Schechtman, K. B., Mulrow, C. D., Fiatarone, M. A., Gerety, M. B., & Buchner, D. M. (1995). Relationship between physical performance and self-perceived physical function. *American Journal of Geriatric Society*, 43, 93-101.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Daltroy, L. H., Larson, M. G., Eaton, H. M., Partridge, A. J., Pless, I. B., Rogers, M. P., Liang, M.H. (1992). Psychosocial adjustment in juvenile arthritis. *Journal of Pediatric Psychology*, 17, 277-289.
- Daltroy, L. H., Logigian, M., Iversen, M. D., & Liang, M. H. (1992). Does Musculoskeletal function deteriorate in a predictable sequence in the elderly? *Arthritis Care Research*, 5, 146-150.
- Deary, I. J. (2000). *Looking Down on Human Intelligence*. Oxford University Press, New York.
- Deary, I. J. (2000). Psychometric intelligence differences and brain function. *Novatis Foundation symposium*, 233, 58-72.
- Deary, I. J. & Derr, G. (2005). Reaction time explains IQ's association with death. *Psychological Science*, 16, 64-69.
- Deary, I. J., & Batty, G. D. (2006). Pre-morbid IQ and later health: the rapidly evolving field of cognitive epidemiology. *International Journal of Epidemiology*. 35, 670-672.
- Dejong, A. & Molenaar, I. W. (1987). An application of Mokken's model for stochastic, cumulative scaling in psychiatric research. *Journal of Psychiatric Research*, 21, 137-149.
- De Morton, N. A., Keating, J. L., & Davidson, M. (2008). Rasch analysis of the Barthel index in the assessment of hospitalized older patients after admission for an acute medical condition. *Archives of Physical Medicine and Rehabilitation*, 89, 641-647.
- Doble, S. E., Fisher, A.G. (1998). The dimensionality and validity of the Older Americans Resources and Services (OARS) Activities of Daily Living (ADL). *Journal of Outcome Measurement Scales*, 2, 4-24.

- Dodge, H.H., Kadowksi, T., Hayakawa, T., Yamakawa, M., Sekikawa, A., & Ueshima, H. (2005). Cognitive impairment as a strong predictor of incident disability in specific ADL-IADL tasks among community-dwelling elders: The Azuchi Study. *The Gerontological Society of America*, 45, 222-230.
- Drachman, D.A., O'Donnell, B.F., Lew, R.A., & Swearer, J.M. (1990). The prognosis in alzheimer's disease. *Archives of Neurology*, 47, 851-856.
- Dubuc, N., Haley, S., Ni, P., Kooyoomjian, J., & Jette, A. (2004). Function and disability in late life: comparison of the Late-Life Function and Disability Instrument to the Short-Form-36 and the London Handicap Scale. *Disability Rehabilitation*, 26, 362-370.
- Dunlop, D. D., Hughes, S.L., Manheim, L. M. (1997). Disability in activities of daily living: Patterns of change and a hierarchy of disability. *American Journal of Public Health*, 87, 378-383.
- Dyck, P. J., Krata K. M., Karnes, J. L. Litchy, W. J., Klein, R., Pach, J.M. Wilson, D.M....Melton, L.J. (1993)The prevalence by staged severity of various types of diabetic neuropathy, retinopathy and nephropathy, retinopathy, and nephropathy in a population based cohort: the Rochester diabetic neuropathy study. *Neurology*, 43, 817-824
- Eisner. M. D., & Iribarren, C. (2007). The influence of cigarette smoking on adult asthma outcomes. *Nicotine & Tobacco Research* 9, 53-6.
- Eisner, M.D., Blanc, P. D, Yelin, E.H., Sidney, S., Katz, P. P, Ackerson, L...Iribarren C.(2008). COPD as a systemic disease: impact on physical functional limitations.*The American Journal of Medicine*,121, 789-96.
- Eisner, M. D., Iribarren, C., Yelin, E. H., Sidney, S., Katz, P., Ackerson, L,...Blanc, P.D..(2008). Pulmonary function and the risk of functional limitation in chronic obstructive pulmonary disease. *American Journal of Epidemiology*, 167, 1090-1101.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8, 341-349.
- Embretson, S. E. (2002). *Generating abstract reasoning items with cognitive theory*. In S. H. Irvine, & P. C. Kyllonen (Eds.), *Item generation for test development*. New Jersey 7 Lawrence Erlbaum Associates.
- Engle, R. W., Tuholski, S. W., Laughlin, J., E; & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128,309-331.

- Ettinger, W. H., Davis, M.A., Neuhaus, J.M., Mallon, K.P. (1994). Long term physical functioning in persons with knee osteoarthritis from NHANES. I: Effects of comorbid medical conditions. *Journal of Clinical Epidemiology*, 47, 809-815.
- Ettingner, A., Reed, M., Cramer, J. (2004). Depression and comorbidity in community-based patients with epilepsy or asthma. *Neurobiology*, 28, 2008-14.
- Fauth, B. E, Zarit, S. H., Malmberg, B., & Johansson, B. (2007). Physical, Cognitive, and Psychosocial Variables From the Disablement Process Model Predict Patterns of Independence and the Transition Into Disability for the Oldest-Old. *The gerontologist*, 47, 613-642.
- Femia, E. E, Zarit, S. H, & Johansson, B. (2001). The disablement process in very late life: a study of the oldest-old in Sweden. *Journal of Gerontology Series B: Psychological and Social Science*, 56, 12-23.
- Ferring, D. and Fillip, S.H. (1995). The structure of subjective well-being in the elderly: A test of different models by structural equation modeling, *European Journal of Psychological Assessment* 11, 32-39
- Festinger, L. (1947). Treatment of Qualitative Data by Scale Analysis. *Psychological Bulletin*, 44, 149-161.
- Figaro M .K, Kritchevsky S. B, Resnick H. E, Shorr R. I, Butler J, Shintani A, Harris, T.B. (2006). Diabetes, inflammation, and functional decline in older adults: findings from the health, aging and body composition study. *Diabetes Care*, 29, 2039-45.
- Fieo, R., Watson, R., Deary, I. J., & Starr, J. M. (2010). A revised activities of daily living/instrumental activities of daily living instrument increases interpretive power: theoretical applications for functional tasks exercise. *Gerontology*, 56, 483-90.
- Fillenbaum, G.G. (1985). Screening the elderly. A brief instrumental activities of daily living measure. *Journal of the American Geriatrics society*, 33, 698-706
- Finlayson, M., Mallinson, T., Barbosa, V. M. (2005). Activities of daily living (ADL) and instrumental activities of daily living (IADL) items were stable over time in a longitudinal study on aging. *Journal of Clinical Epidemiology*, 58, 338-349.
- Finch, M, Kane, R. L., & Philip, I. (1994). Developing a new metric for ADLs. *Journal of American Geriatric Society*, 43, 877-884.

- Finkel, S. I., Mintzer, J. E., Dysken, M., Krishnan, K. R., Burt, T., & McRae, T. (2004). A randomized, placebo-controlled study for the efficacy and safety of sertaline in the treatment of the behavioural manifestations of Alzheimer's disease in outpatients treated with donepezil. *International Journal of Geriatric Psychiatry*, 19, 9-18.
- Fisher, W. P., & Fisher, A.G. (1993). Applications of Rasch analysis to studies in occupational therapy. *Physical Medicine and Rehabilitation Clinics of North America* 4, 551-569.
- Fisher, A. G. (1992). The assessment of IADL motor skills: an application of many-faceted Rasch analysis. *American Journal of Occupational Therapy* , 47, 319-329.
- Fiske, A., Wetherell, J.L., & Gatz, M. (2009). Depression in older adults. *Annual Review of Clinical Psychology*, 53, 63-89.
- Fitti, J. E., & Kovar, M.G. (1987). The Supplement on Aging to the 1984 National Health Interview Survey. *Vital Health station*, 21, 1-115.
- Fitzpatrick, R., Davey, C., Buxton, M. J., & Jones D. R. (1998).Evaluating patient-based outcome measures for use inclinical trials. *Health Technology Assessment* 2 (14).
- Fleishman, J. A., & Spector, W. D. (2002). Altman BM. Impact of differential item functioning on age and gender differences in functional disability. *Journal of Gerontology Series B: Psychological Sciences and Social Sciences*, 57, 275-284.
- Forrest, Y. Z., Zmuda, J.M., & Cauley, J.A. (2007). Patterns and correlates of muscle strength loss in older woman. *Gerontology*, 53, 140-147.
- Fortinsky, R. H., Garcia, R. I., Sheehan, J.T., Madigan, E. A., & Tullai-McGuinness, S. (2003). Measuring disability in Medicare home care patients: application of Rasch modelling to outcome and assessment information set. *Medical Care Research and Review*, 41, 601-615.
- Foubert-Samier A, Catheline, G., Amieva, H., Dilharreguy, B., Helmer, C., Allard, M., Dartigues, J.F. (2010). Education, occupation, leisure activities, and brain reserve: a population-based study. *Neurobiology of ageing*, Nov 11.
- Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, 78, 350-365.

- Freedman, V.A., Martin, L.G., Schoeni, R.F. (2002). Recent trends in disability and functioning among older adults in the United States: a systematic review. *Journal of the American Medical Association*, 288, 3137-3146.
- Fried, L. P., S. J. Herdman, Rubin, G., & Turano, K. (1991). Preclinical disability: hypotheses about the bottom of the iceberg. *Journal of Ageing Health*, 3, 285-300.
- Fried, L. P., W. H. Ettinger, Newman, A. B., & Gardin, J. (1994). Physical disability in older adults: A physiological approach. *Journal of Clinical Epidemiology*, 47,747-760.
- Fried, L. P., Brandeen-Roche, K., Williamson, J. D., Prasada-Rao, P., Chee, E., Tepper, S., & Rubin, G.S. (1996). Functional decline in older adults: expanding methods of ascertainment. *The Journals of Gerontology, Series A: A Biological Science*, 51, 206-214.
- Fried, L. P., & Guralnik. J. M. (1997). Disability and older adults: Evidence regarding significance, etiology and risk. *Journal of the American Geriatric and Journal of the American Medical Association*, 276, 1473-1479.
- Fried, L. P., Kronmal, R. A., Newman, A. B., Bild, D. E., Mittelman, M. B., Polak, J. F., & Gardin, J. M.(1998). Risk factors for 5-year mortality in older adults: The Cardiovascular Health Study. *JAMA*, 279, 585–592.
- Fried, L.P., Bandeen-Roche, K., Chaves, P.H., & Johnson, B. (2000). Preclinical mobility disability predicts incident mobility disability in older women. *Journal of Gerontological Medical Science*, 55, 43-52.
- Fried, L.P, Bernardini, J., & Piraino, B. (2003). Comparison of the Charlson Comorbidity Index and the Davies score as a prediction of outcomes in PD patients. *Peritoneal Dialysis International*, 23, 568–573.
- Fried, L.P.(2003).Establishing Benchmarks for Quality Care for an Aging Population. *Caring for Vulnerable Older Adult*, 139, 784-786
- Fried, L. P., Ferrucci, L., Darer, J., Williamson, J. D., & Anderson, G. (2004) Untangling the concepts of disability, frailty, and comorbidity:Implications for improved targeting and care. *Journal of Gerontology: Medical Sciences*, 59, 255–263.
- Fried, L. F., Katz, R., Sarnak, M. J., Shlipak, M. G., Chaves, P. H., Jenny, N.S., Newmanet, A.B. (2005). Kidney function as a predictor of noncardiovascular mortality. *Journal of the American Society of Nephrology*,16, 3728–3735.

- Fried, T.R., van, D.C., O'Leary, J.R., Tinetti, M.E., & Drickamer, M.A. (1999). Older persons' preferences for site of terminal care. *Annals of Internal Medicine* 131, 109-112.
- Fried, T.R., Bradley, E.H., O'Leary, J., (2003). Prognosis communication in serious illness: perceptions of older patients, caregivers, and clinicians. *Journal of the American Geriatrics Society*, 51, 1398-403.
- Fries, J.F., Bioch, D.A., Harrington, H., Richardson, N., Beck, R. (1993) Two-year results of a randomized controlled trial of a health promotion program in a retiree population: the Bank of America Study. *American Journal of Medicine*, 94, 455-62.
- Fries, J.F., Harrington, H., Edward, R., Kent, L.A., Richardson, N. (1994). Randomized controlled trial of cost reductions from a health education program: the California Public Employees' Retirement System (PERS) study. *American Journal of Health Promotion*, 8, 216-23.
- Fries, J.F. (2003). Measuring success in Compressing Morbidity. *Annals of Internal Medicine*, 139, 455-459.
- Fries, J.F., Carey, C., McShane, D. J. (1997). Patient education in arthritis: randomized controlled trial of a mail-delivered program, *Journal of Rheumatology*, 24, 1378-1383.
- Fries, J. F. (2006). The promise of the future, updated: Better outcome tools, greater relevance, more efficient study, lower research costs. *Future Rheumatology*, 1, 415-21.
- Fries, J. F., Bruce, B., Bjorner, J., & Rose, M. (2006). More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. *Annals of the Rheumatic Diseases*, 65, 16-21.
- Gardner, M. M., Buchner, D.M., Robertson, M.C., & Campbell, A.J. (2001). Practical implementation of an exercise-based falls prevention programme. *Age Ageing*, 30, 77-83.
- Gardner, E. A, Huppert, F.A, & Guralnik, J.M.(2006).Middle-aged and mobility-limited, prevalence of disability and symptom attributions in a national survey. *Journal of General International Medicine*, 21, 1091–1096.
- Gardener, L. & Melzer, D. (2005), Mobility Disability Self-Reporting and Gait Speed in England and the USA, University of Cambridge and Peninsula Medical School, http://www.ifs.org.uk/elsa/publications.php?publication_id=3451. Accessed 8 June 2006.
- Griffiths, T.L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J.B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14, 357–364.

- Grimby, G., & Saltin B(1983). The ageing muscle. *Clinical Physiology*, 3, 209-218,
- Gill, T. M, Baker, D. I, Gottschalk, M., Peduzzi, P.N., Allore, H. & Byers, A.(2002). A program to prevent functional decline in physically frail, elderly persons who live at home. *Nationl English Journal of Medicine*, 347,1068-1074.
- Gevins, A. & Smith, M. E. (2000). Neuropsychological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, 10, 829-839.
- Goel, V. Gold, B. Kapur, S. & Houle, S. (1997). The seats of reason? An imaging study of deductive and inductive reasoning. *Neurological Report*, 8,1305-1310.
- Godin, O., Dufouil, C., Maillard, P., Delcroix, N., Mazoyer, B., Crivello, F.,...Tzourio, C. (2008). White Matter Lesions as a Predictor of Depression in the Elderly: The 3C-Dijon Study. *Biological Psychiatry*, 63, 663-669.
- van Gool, C.H., Kempen, G.I., Penninx, B. W. Deeg, D.J., Beekman, A.T. & van Eijk, J. T. (2003)Relationship between changes in depressive symptoms and unhealthy lifestyles in late middle aged and older persons: results from the Longitudinal Aging Study Amsterdam. *Age and Ageing*, 32, 81-7.
- Gow, A. J., Johnson, W., Pattie, A. Whiteman, M.C., Starr, J., & Deary, I.J. (2008). Mental ability in childhood and cognitive aging. *Gerontology*, 54, 177-86.
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, 6, 316-322.
- Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: science and ethics *Nature Reviews Neuroscience*, 4, 471-482.
- Griffith, L., Raine, P., Wu, H., Zhu, B., & Stathothostas, L. (2010). Population Attributable Risk for Functional Disability Associated with Chronic Conditions in Canadian Older Adults. *Age and Ageing*, 39, 738-745.
- Gross, J. J., Carstensen, L. L., Pasupathi, M., Tsai, J., Götestam S. C., & Hsu, A. Y. C. (1997). Emotion and aging: experience, expression, and control. *Psychology of Aging*, 12, 590–599.

- Gatz, M., Svedberg, P., Pedersen, N. L., Mortimer, J. A., Berg, S., & Johansson, B. (2001). Education and the risk of Alzheimer's disease: Findings from the Study of Dementia in Swedish Twins. *Journal of Gerontology: Psychological and Social Sciences*, 56, 292-300.
- Giampoli, S., Ferrucci, L., Cecchi, F., Noce, C.L., Poce, A., Dima, F., Santaquilani, A., Vescio, M.F., & Menotti, A. (1999). Hand-grip strength predicts incident disability in non-disabled older men. *Age and Aging*, 28, 283-288.
- Gibbons, R. D., Clark, D. C., Cavanaugh, S. V., & Davis, J. M. (1985). Application of modern psychometric theory in psychiatric research. *Journal of Psychiatric Research*, 19, 43-55.
- Gill, T. M., Baker, D. I., Gottschalk, M., Peduzzi, P. N., Allore, H., & Byers, A. (2002). A program to prevent functional decline in physically frail, elderly persons who live at home. *The New England Journal of Medicine*, 347, 1068-74.
- Gill, T. M., Han, L., & Allore, H. G. (2007). Predisposing factors and precipitants for bathing disability in older persons. *Journal of American Psychiatric Society*, 55, 534-540.
- Gillespie, M., Tenvergert, E. M., & Kingma, J. (1987). Using Mokken scale analysis to develop unidimensional scales. *Quality and Quantity*, 21, 393-408.
- Giovannetti, T., Libon, D. , & Hart, T. (2002). Awareness of naturalistic action errors in dementia. *Journal of International Neuropsychological Society*, 8, 633-644.
- Glantz, S.A. (1992). *Primer of Biostatistics*. 3rd ed. McGraw Hill, New York.
- Gottfredson, L.S., & Deary, I.J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychology and Science*, 13, 1-5.
- Grundy, E. & Glaser, K. (2000). Socio-demographic differences in the onset and progression of disability in early old age: a longitudinal study, *Age Ageing*, 29, 149-157.
- Grundy, E. & Holt, G. (2001). The socio-economic status of older adults: how should we measure it in studies of health inequalities ? *Journal of Epidemiology and Community Health*, 55, 895-904.
- Guimarães, R. M. (2007). Health capital, life course and ageing. *Gerontology*, 53, 96- 101.
- Gunning-Dixon, F. & Raz, N. (2003). Neuroanatomical correlates of selected executive functions in middle-aged and older adults: a prospective MRI study. *Neuropsychologia*, 41, 1929-41.

- Guralnick, J. M., Branch, L. G., Cummings, S. R., & Curb, J. D. (1989). Physical performance measures in aging research. *Journal of Gerontology*, 44, 141-146.
- Guralnick, J.M., & Simonsick, E. M. (1993). Physical disability in older Americans. *Journal of Gerontology*, 48, 3–10.
- Guralnik, J.M., La Croix, A.Z., Abbott, R.D., Berkman, L.F., Satterfield, S., Evans, D.A., & Wallace, R.B. (1993). Maintaining mobility in late life. I. Demographic characteristics and chronic conditions. *American journal of Epidemiology*, 137, 845-857.
- Guralnik, J. M., Simonsick, E. M., Ferrucci, L., Glynn, R. J., Berkman, L. F., Blazer, D. G., Scherr, P. A. & Wallace, R. B. (1994). A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 49, 85–M94.
- Guralnik, J.M. (1994). Understanding between disease and disability (ed). *Journal of the American Geriatrics Society*, 42, 1128-1129.
- Guralnik, J.M., Ferrucci, L., Simonsick, E.M., Salive, M., & Wallace, R. (1995). Lower-extremity functioning in persons over the age of 70 years as a predictor of subsequent disability. *National English Journal of Medicine*, 332, 556-561.
- Guralnik, J.M., Ferrucci, L., Balfour, J.L., Volpato, S., & Di Lorio, A. (2001). Progressive versus catastrophic loss of the ability to walk: implications for the prevention of mobility loss. *Journal of American Geriatric Society*, 49, 1463–1470.
- Guralnik, J. M., & Ferrucci, L. (2003). Assessing the building blocks of function: utilizing measures of functional limitation. *American Journal of Preventive Medicine*, 25, 112–121.
- Guralnik, J.M., & Ferrucci, L. (2009). Demography and epidemiology. In: J.B. Halter, J.G. Ouslander, M.E. Tinetti et al. (eds) *Hazzard's Geriatric Medicine and Gerontology*, 6th edn. McGraw-Hill: New York.
- Guralnik J.M. & Ferrucci, L. (2009). The challenge of understanding the disablement process in older persons; commentary responding to Jette AM. Toward a common language of disablement. *Journal of Gerontology Series A: Biological Science and Medical Science*, 64A, 1169-1171. doi:10.1093/Gerona/glp094.

- Gustafsson, J. E. (1988). *Hierarchical models of individual differences in cognitive abilities*. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. 4, 35–71. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Guttman, L.A. (1944). A basis for scaling qualitative data. *American Sociological Review*, 91, 139-150.
- Guttman, L. (1950). *The basis for scalogram analysis*. In *Measurement and prediction*. Edited by Stouffer SA & et al. New York: Wiley; 1950:60-90.
- Guyote, M. J. & Sternberg, R. J. (2003). A transitive-chain theory of syllogistic reasoning, *Cognitive Psychology*, 13, 461–525.
- Hajjar, I., Yang, F., Sorond, F., Jones, R.N., Milberg, W., Cupples, L.A., Lipsitz, L.A. (2009). A novel aging phenotype of slow gait, impaired executive function, and depressive symptoms: Relationship to blood pressure and other cardiovascular risks. *J Gerontol A Biol Sci Med Sci*, 64, 994-1001.
- Haley, S. M., McHorney, C.A., & Ware, J.E. (1994). Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *Journal of Clinical Epidemiology*, 47, 671-684.
- Haley, S. .M., Fragala, M. A., Aseltine, R., Ni, P., & Skrinar, A. (2003). Development of a disease-specific disability instrument for Pompe disease. *Pediatric Rehabilitation*, 6, 77-84.
- Haley, S. M., Coster, W. J., Andres, P, L., Ludlow, L. H., Ni, P., Bond, T. L., Sinclair, S. J, & Jette, A. M. (2004). Activity outcome measurement for postacute care. *Medical Care*, 42, 49-61.
- Hall, P. A., Dubin, J. A., Crossley, M., Holmqvist, M. E., & D’Arcy, C. (2009). Does executive function explain the IQ-mortality association? Evidence from the Canadian study on health and aging. *Psychosomatic Medicine*, 71, 196–204.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.
- Harvey, R. J., & Hammer, A. L. (1999). Item response theory. *Counselling Psychology*, 27, 353-383.
- Hattie, J. A. (1985). Methodological review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139-164.

- Hemmingsson, T., Melin, B., Allebeck, P., Lundberg, I.(2006). The association between cognitive ability measured at ages 18–20 and mortality during 30 years of follow-up: a prospective observational study among Swedish males born 1949–1951. *International Journal of Epidemiology*, 35, 665–70.
- Hoeymans, N., Feskens, J.M., Geertudis, A.M., & Kromhout, D. (1997). Age, time, and cohort effects on functional status and self-rated health in elderly men. *American Journal of Public Health*, 87, 1620–1625.
- Hoffman, C., Rice, D. P., & Sung, H. Y. (1996). Persons with chronic conditions: *Their prevalence and costs*. *Journal of the American Medical Association*, 276, 1473-1479.
- Hopman-Rock, M., Van Buuren, S., & De Kleijn-De Vrankrijker, M.(2000). Polytomous Rasch analysis as a tool for revision of the severity of disability code of the ICIDH. *Journal of Disability and Rehabilitation*, 20,363-71.
- Hoyer, W. J. Stawski, R.S. Wasylshyn C. & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis, *Psychology and Aging* 19, 211–214.
- Hoeymans, N., Feskens, E. J., Van den Bos, G. A., & Kromhout, D. (1996). Measuring functional status: cross-sectional and longitudinal associations between performance and self-report. *Journal of Clinical Epidemiology*, 49, 1103-1010.
- Hortobagyi, T., Mizelle, C. Beam, S., & De Vita, P. (2003). Old adults perform activities of daily living near their maximal capabilities. *Journal of Gerontology:Medical Sciences*, 58, 453-460.
- Hulin, L. H., Drasgow, Y., & Parsons, C. K. (1983). *Item response theory: application to psychological measurement*. Homewood IL:Dow Jones-Irvin.
- Hurst, N. P. (1994). An evaluation of the Health Assessment Questionnaire (HAQ) in a long-term longitudinal follow-up of disability in rheumatoid arthritis. *British Journal of Rheumatology*, 33,196-1997.
- Hybels. C.F. & Blazer, D.G. (2003). Epidemiology of late-life mental disorders. *Clinics in Geriatric Medicine*, 19, 663-696.
- Institute of medicine. (2003).*Priority areas for national action: transforming health care quality*. New York: national academy Press.

- Institute of Medicine. (2003). Who will keep the public healthy? Educating public health professionals for the 21st century. Gebbe K, Rosenstock K, & Hernandez LM, Eds. Washington, DC: National Academy Press.
- Inouye, S. K., Peduzzi, P. N., Robison, J. T., Hughes, J. S., Horwitz, R. I., & Concato, J. (1998). Importance of functional measures in predicting mortality among older hospitalized patients. *Journal of American Medical Association*, 279, 1187-1193.
- Jackson, R. J., Draugalis, J. R., Slack, M. K., Zachry, W. M., & D'Agostino, J. (2002). Validation of authentic performance assessment: a process suited for Rasch modelling. *American Journal of Pharmacological Education*, 66, 233-243.
- Jagger, C., Matthews, R., Melzer, D., Matthews, F., & Brayne, C. (2007). Educational differences in the dynamics of disability incidence, recovery and mortality: findings from the MRC cognitive function and ageing study (MRC CFAS). *International Journal of Epidemiology*, 36, 358-65.
- Jefferson, A.L., Paul, R.H., Ozonoff, A. & Cohen, R.A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living (IADLs). *Archives of Clinical Neuropsychology*, 21, 311-320.
- Jette, A. M., Haley, S. M., Coster, W. J., Kooyoomjian, J. T., Levenson, S., Heeren, T., & Ashba, J. (2002). Late life function and disability instrument: I. Development and evaluation of the disability component. *The Journals of Gerontology, Series A: A Biological Science*, 57, 209-216.
- Jensen, A. R. (1998). *The g factor; The science of mental ability*. Westport, CT, us: Praeger Publishers/Greenwood Publishing Group, Inc.
- Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Science*, 5, 434-442.
- Joynt, K. E., Whellan, D. J., & O'Connor, C. M. (2003). Depression and cardiovascular disease: mechanisms of interaction. *Biological Psychiatry*, 54, 248-261.
- Joy, S., Fein, D., Kaplan, E., & Freedman, M. (2000). Speed and memory in WAIS-R-NI Digit Symbol performance among healthy older adults, *Journal of the International Neuropsychology Society*, 6, 770-780.
- Joy, S., & Fein, D. (2001). Meta-analysis: Aging, Digit Symbol, and Symbol Copy. Poster presented at the annual convention of the Eastern Psychological Association, Washington, DC.

- Joy, S., Kaplan, E., & Fein, D., (2004). Speed and memory in the WAIS-III Digit Symbol—Coding subtest across the adult lifespan. *Archives of Clinical Neuropsychology*, 19, 759–767.
- Junker B.W., & Sijtsma K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Junker, B.W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (p. 247-276). New York: Springer-Verlag.
- Kallman, D. A., Plato, C. C. & Tobin, J. D. (1990). The role of muscle loss in the age-related decline of grip strength: cross-sectional and longitudinal perspectives. *J. Gerontol.* 45: M82–M88.
- Kane, M. J. & Engle, R. W. (2002). The role of pre-frontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychological Bulletin Review*, 9, 637-671.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs : Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66-71.
- Kanda, A. Matsui, T. Ebihara, S. Arai, H. & Sasaki H. (2003). Periventricular white matter lesions and sleep alteration in older people, *American Journal of Geriatric Society*, 51, 432–433.
- Kattainen, A., Reunanen, A., Koskinen, S., & Martelin, T. (2004). Disability predicted mortality in men but not women with coronary heart disease. *Journal of Clinical Epidemiology*, 57, 513-521
- Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., & Jaffe, M. W. (1963). Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, 185, 914-919.
- Katzman, R. (1993). Education and the prevalence of dementia and Alzheimer's disease. *Neurology* 43, 13-20.
- Keisuke, S., Koujiro, K., Tomoaki, S. (2006). Longitudinal relationships between ability to perform activities of daily living and mobility in community-living elderly over 75 years old. *Bulletin of health Sciences Kobe*, 21, 13-25.
- Kempen, G. M., & Suurmeijer, T. P. (1990). The development of a hierarchical polychotomous ADL-IADL scale for noninstitutionalized elders. *Gerontologist*, 30, 407-502.

- Kempen, G.I., Brilman, E. L., Heyink, J. W., & Ormel, J. (1995). *MOS Short-Form General Health Survey (SF-20)* Noordelijk centrum voor gezondheidsvraagstukken: Groningen.
- Kempen, G. I., Myers, A., & Powell, L. (1995). Hierarchical structures in ADL and IADL: analytical assumptions and applications for clinicians and researchers. *Journal of Clinical Epidemiology*, 48, 1299-1305.
- Kempen, G. M., Miedema, I., Ormel, J., & Molenaar, W. (1996). The assessment of disability with the Groningen activity restriction scale. Conceptual framework and psychometric properties. *Journal of Social Science and Medicine*, 43, 1601-1610.
- Kempen, G. I., Stevernik, N., Ormel, J., & Deeg, D. J. (1996). The assessment of ADL among frail elderly in an interview survey: Self-report versus performance-based tests and determinants of discrepancies. *Journal of Gerontology: Psychological Sciences*, 51, 254-260.
- Kempen, G.I., & Ormel, J. (1998). The impact of physical performance and cognitive status on subsequent ADL disability in low-functioning older adults. *International Journal of Geriatric Psychiatry*, 13, 480-3.
- Kempen G.M., Verbrugge, L. M., Merrill, S. S., & Ormel, J. (1998). The impact of multiple impairments on disability in communitydwelling older people. *Age and Ageing*, 27, 595-604.
- Kempen, G. I., Sullivan, M., Van Sonderen, E., & Ormel, J. (1999). Performance-based and self-reported physical functioning in low-functioning older persons: Congruence of change and the impact of depressive symptoms. *Journal of Gerontology –Series B*, 54, 380-386.
- Kelly-Hayes, M. (1992). Functional limitations and disability among elders in the Framingham study. *American Journal of Public Health*, 82, 841-845.
- Khokhar, Y. Stern, K. Bell, K. Anderson, E. Noe, R. Mayeux, S.M. & Albert, S.M., (2001). Persistent mobility deficit in the absence of deficits in activities of daily living: a risk factor for mortality. *Journal of American Geriatric Society*, 49, 1539–1543.
- Kinugasa, T., & Nagasaki, H. (1998). Reliability and validity of the motor fitness scale for older adults in the community. *Aging*, 10, 295-302.

- Koopmans, G. T., & Lamers, L. M. (2001). Comparing patients with depressive complaints and patients with chronic medical conditions on their functioning and medical consumption. *Journal of Mental Health Policy Economy*, 4, 91–100.
- Koopmans, G.T. & Lamers, L. M. (2006). Is the impact of depressive complaints on the use of general health care services dependent on severity of somatic morbidity? *Journal of Psychosomatic Research*, 61, 41-50.
- Koyano, W., Shibata, H., Nakazato, K., Haga, H., Suyama, Y. & Matsuzaki, T. (1988). Prevalence of disability in instrumental activities of daily living among elderly Japanese. *Journal of Gerontology Social Science*, 43, 41-45.
- Kriegsman, D. M., van Eijk, J. T., Penninx, B. W., Deeg, D. J. H. & Boeke, A. J. (1997). Does family support buffer the impact of specific chronic diseases on mobility in community dwelling elderly? *Disability & Rehabilitation*, 19, 71-83.
- Lamberts S. W. J., van den Beld A. W., & van der Lely A.J. (1997). *The endocrinology of aging*, Science, 278, 419-424.
- LaPlante, M. P. (2010). The classic measure of disability in activities of daily living is biased by age but an expanded IADL/ADL measure is not. *Journal of Gerontology Series B: Psychological and Social Science*, 65, 720-732.
- Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, 9, 179-186.
- Lawton, M. P., Kleban, M. H., Rajagopal, D., & Dean, J. (1992). Dimensions of affective experience in three age groups, *Psychology and Aging*, 7, 171–184.
- Lauretani, F., Russo, C. R., Bandinelli, S., Bartali, B., Cavazzini, C., Angelo Di Iorio, A.D.,...Ferrucci, L. (2003). Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia.. *Journal of Applied physiology*, 95, 1851-186.
- Lazaridis, E. N., Rudberg, M. A., Furner, S. E., & Cassel, C. K. (1994). Do activities of daily living have a hierarchical structure? An analysis using the longitudinal study of aging. *Journal of Gerontology*, 49, 47-51.
- Lee, K.H, Choi Y.Y., Gray, J.R., Cho, S. H., Chae, J., Lee, S., & Kim, K.(2006). Neural correlates of superior intelligence: stronger recruitment of posterior parietal cortex. *Neuroimage*, 29, 578-586.

- Lenze, E. J., Mulsant, B. H., Mohlman, J. Shear, K., Dew, M.A., Schulz, R., Miller, M.D., Tracey, B., & Reynolds, C. F. (2005). Generalized anxiety disorder in late life. *American Association for Geriatric Psychiatry, 13*, 77-80.
- Lessov-Schlaggar, C. N, Swan, G. E, Reed, T, Wolf, P. A, & Carmelli, D. (2007). Longitudinal genetic analysis of executive function in elderly men. *Neurobiology and Aging, 28*, 1759–68.
- Leveille, S.G., Ling, S., Hochberg, M.C., Resnick, H.E., Bandeen-Roche, K.J., Won, A., & Guralnik, J.M. (2001). Widespread musculoskeletal pain and the progression of disability in older disabled women. *Annual International Medicine, 135*, 1038–1046.
- Liao, P. M., & Campbell, S. K. (2004). Examination of the item structure of the Alberta infant motor scale. *Pediatric Physical Therapy, 16*, 31-38.
- Ligtvoet, R., van der Ark, L. A., te Marvelde, J.M., & Sijtsma, K. (2010). Investigating an invariant item ordering for polytomously scored items. *Educational Psychological Measurement, 40*, 578–595.
- Lindmark, B., & Hamrin, E. (1995). A five-year follow-up of stroke survivors: Motor function and activities of daily living. *Clinical Rehabilitation, 9*, 1–9.
- Loevinger, J. (1948). The technique of homogenous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin, 45*, 507-29.
- Lohman, D.F. (2001). *Fluid intelligence, inductive reasoning, and working memory: Where the theory of Multiple Intelligences falls short*. The University of Iowa. IN.
- Lohman D. F. & Joni M. Lakin. (2009). *Reasoning and Intelligence*. The University of Iowa. To appear in R. J. Sternberg and S. B. Kaufman (Eds.), *Handbook of Intelligence* (2nd ed.). New York: Cambridge University.
- Lorig K.R., Sobel D.S., Stewart A.L., Brown Jr. B.W., Ritter P.L., González V.M., Laurent D.D., & Holman H.R. (1999). Evidence suggesting that a chronic disease self-management program can improve health status while reducing utilization and costs: A randomized trial. *Medical Care, 37*, 5-14.
- Louie, G. H, & Ward, M. M. (2010). Association of measured physical performance and demographic and health characteristics with self-reported limitations. *Health & Quality of Life Outcomes, 8*, 1-13.

- Lundgren-Nilsson, A., Grimby, G., Ring, H., Tesio, L., Lawton, G., Slade, A., Penta, M.,...Tennant A. (2005). Cross-cultural validity of functional independence measure items in stroke: a study using Rasch analysis. *Journal of Rehabilitation Medicine*, 37, 23-31.
- Lynch N. A., Metter E. J., Lindle R. S., Fozard J. L., Tobin J. D., Roy T. A., Fleg J. L., & Hurley B. F.(1999) Muscle quality. I. Age-associated differences between arm and leg muscle groups. *Journal of Applied Physiology*, 86,188–194.
- Manton, K. G., & X. Gu. (2001). Changes in the prevalence of chronic disability in the United States black and nonblack population above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences*, 98, 6354-6359.
- Marchionni, N., Bar, M. D. fumagalli, S., Ferrucci, L., Baldereschi, G., Timpanelli, M., & Giulio, M.(1996) Variable effect of comorbidity on the association of chronic cardiac failure with disability in community-dwelling older persons. *Archives of Gerontology and Geriatrics*, 23, 283-292.
- Martin, M., Kosinski, M., Bjorner, J.B., Ware, J.E., & MacLean, L.T. (2007). Item response theory methods can improve the measurement of physical function by combining the modified health assessment questionnaire and the SF-36 physical function scale. *Quality of Life Research*, 16, 647–660.
- Matthews, R. J., Jagger, C. & Hancock, R.M. (2005). Does socio-economic advantage lead to a longer, healthier old age? *Social Science & Medicine*, 62, 2489-2499.
- May, K., & Nicewander, W. A. (1998). Measuring change conventionally and adaptively. *Educational Psychological Measurement*, 58, 882-897.
- McAuley, E., Konopack J.F., Motl, R.W., Rosengren, K., & Morris, K.S. Measuring disability and function in older women: psychometric properties of the late-life function and disability instrument. *Journal of Gerontology: MEDICAL SCIENCES*, 60A, 901-909.
- McGee, M. A, Johnson, A. L, & David, W. K. (1998). The description of activities of daily living in five centres in England and Wales. *Age and Aging*, 27,605-613.
- McHorney, C. A., Haley, S. M., & Ware, J. E. (1987). Evaluation of the MOS SF-36 physical functioning scale (PF-10) II: Comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, 50,451–461.
- McHorney, C. A., & Tarlov, A. R. (1996). Measuring and monitoring general health status in elderly persons: practical and methodological issues in using the SF-36 Health Survey. *Gerontologist*, 5, 571-583.

- McHorney, C. A., Haley, S. M., & Ware, J. E. (1997). Evaluation of the MOS SF-36 physical functioning scale (PF-10) II: comparison of relative precision using Likert and Rasch scoring methods. *Journal of Clinical Epidemiology*, 50, 451–461.
- McHorney, C. A., & Cohen, A. S. (2000). Equating health status measures with item response theory: illustrations with functional status items. *Medical Care*, 38, 43-59.
- McHorney, C. A. (2002). Use of item response theory to link 3 modules of functional status items from the asset and health dynamics among the oldest old study. *Archives of Physical Medicine and Rehabilitation*, 83, 383-394.
- Meijer, R.R., Sijtsma, K. Y., & Smid, N.G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement*, 14, 283-298.
- Meguro, K., Shimada, M., Yamaguchi, S., Ishiizaki, J., Yamadori, A., & Sekita, Y. (2001). Five-year interval changes of cognitive screening tests performances in the elderly: a community-based study. *Journal of Gerontology: Psychological Science*, 56, 314-318.
- Melzer, D., Gardener, E., & Guralnik, J. M. (2005). Mobility disability in the middle-aged: cross-sectional associations in the English Longitudinal Study of Ageing. *Age and Ageing*, 34, 594-602.
- Melzer, D., McWilliams, B., Brayne, C., Johnson, T., & Bond, J. (2000). Socioeconomic status and the expectation of disability in old age: estimates for England. *Journal of Epidemiology & Community Health*, 54, 286–92.
- Menzel, H. (1953). A new coefficient for Scalogram analysis. *Public Opinion Quarterly*, 17, 268-280.
- Messick, S. (1989). Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not be equal to the sum of the parts. *Journal of Clinical Epidemiology*, 49, 711-717.
- Miravittles M, Anzueto A. (2009). Insights into interventions in managing COPD patients: Lessons from the TORCH and UPLIFT (registered trademark) studies *International Journal of COPD*, 4, 185-201.
- Mokken, R.J. (1971). *A Theory and procedure of scale analysis*. The Hague: Mouton.
- Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417-430.

- Morris, J.N., Fries, B.E., & Morris, S.A. (1999). Scaling ADLs within the MDS. *Journal of Gerontology: MEDICAL SCIENCES*, 54, 546-553.
- Mokken, R. J. & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6, 417-430.
- Molenaar, I. W. & Sijtsma, K.,(2000) User's manual MSP5 for Windows, Groningen: iec ProGAMMA., p105.
- Moorer, P., & Suurmeijer, T. (1994). A study of the unidimensionality and cumulativeness of the MOS short-form general health survey. *Psychological Report* , 74, 467–470.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14, 59-71.
- Nandakumar, R. (1994) Assessing dimensionality of a set of item responses – comparison of different approaches. *Journal of Educational Measurement*, 31:17–35.
- Nagi, S.Z. (1964) A study in the evaluation of disability and rehabilitation potential: concepts, methods, and procedures. *American Journal of Public Health*. ,54,1568–1579.
- Nagi, S. Z. (1965). Some conceptual issues in disability and rehabilitation. In *Sociology and Rehabilitation*. Edited by Sussman MB. Washington, DC: American Sociological Association; 100-113.
- Nagi, S.Z. (1976). An epidemiology of disability in the United States. *Milbank Memorial.Fund Quarterly Health and Society*, 54, 439-467.
- Naik, A. D., Concato, J., & Gill, T. M. (2004). Bathing disability in community-living older persons: common, consequential, and complex. *Journal of American Psychiatric Society*, 52, 1805–1810.
- Ng, T., Niti, M., Chiam, P., & Kua, E. (2006) Physical and cognitive domains of the instrumental activities of daily living: validation in a multiethnic population of Asian older adults. *The Journals of Gerontology, Series A: A Biological Science*, 61,726-735.
- Nishiwaki, T., Nakamura, K., Ueno, K., Fujino, K. & Yamamoto, M. (2005). Health characteristics of elderly Japanese requiring care at home. *Tohoku Journal of Experimental Medicine*, 205, 231-239.
- Norstrom, T. & Thorslund, M. (1991). The structure of IADL and ADL measures: some findings from a Swedish study. *Age and Ageing*, 20, 23-28.

- Orme, J., G., Reis, J. and Herz, E., J., (1986) Factorial and discriminant validity of the Center for Epidemiologic Studies Depression (CES-D) scale. *Journal of Clinical Psychology*, 42, 28–33
- O'Toole, B., & Stankov, L. (1992). Ultimate validity of psychological tests. *Personality and Individual Differences*, 13, 699-716.
- Park J., Y., Chong, A. Y., Cochran, E., K., Kleiner, D., E., Haller, M., J., Schatz, D., A., and Gorden, P., G., (2007). Type 1 Diabetes Associated with Acquired Generalized Lipodystrophy and Insulin Resistance: The Effect of Long-Term Leptin Therapy. *The Journal of Clinical Endocrinology & Metabolism*, 10, 1210-1856.
- Peek, K. M., Ottenbacher, K. J. Markides, K. S., & Ostir, G. V. (2003). Examining the disablement process among older Mexican American adults. *Social Science and Medicine*, 57, 413-425.
- Penninx, B. W. J. H., Geerlings, S. W., & Deeg, D. J. H. (1999) Minor and major depression and the risk of death in older persons. *Archives of General Psychiatry*, 56, 889-895
- Peres, K., Verret, C., Alioum, A., & Barberger-Gateau, P. (2005). The disablement process: Factors associated with progression and disability and recovery in French elderly people. *Disability and Rehabilitation*, 27, 263–276.
- Penninx, B. W. J. H., Guralnik J. M., Ferrucci, L., Simonsick, E.M., Deeg, D.J., & Wallace, R.B. (1998) Depressive symptoms and physical decline in community-dwelling older persons. *Journal of the American Medical Association*, 279, 1720-1726.
- Perfetti, B., Saggino, A., Ferretti, A., Caulo, M., Romani, G. L., Onofri, M. (2009). Differential patterns of cortical activation as a function of fluid reasoning complexity. *Human Brain Mapping*, 30, 497-510.
- Perls, T. (2004). The New England Centenarian Study. *Experimental Gerontology*, 39, 1587-93.
- Phillips, W.T. & Haskell, W.L. (1995). Muscular fitness: easing the burden of disability for elderly adults. *Journal of Aging and Physical Activity*, 3, 261-289.
- Phelan J, Link BG, Diez-Roux A, Kawachi I, Levin B (2004) "Fundamental Causes" of Social Inequalities in Mortality: A Test of the Theory. *Journal of Health and Social Behavior*, 45: 265-285.
- Plichart, M., Barberger-Gateau, P., Tzourio, C., Amouye, P., Pérès, K., Ritchie, K., & Empana, J., P. (2010) Disability and Incident Coronary Heart Disease in Older Community-Dwelling Adults: The Three-City Study. *Journal of the American Geriatric Society*, 58, 636-42.

- Primi, R. (2000). Complexity of geometric inductive reasoning tasks Contribution to the understanding of fluid intelligence. *Intelligence*, 29, 41-70.
- Raina, P., Dukeshire, S., Chambers, L. W., Denton M. & Gafni A. (1998) Income, Disability and Functional Independence. ISEOP Research Paper (in press): Faculty of Social Sciences, McMaster University, Hamilton, Ontario.
- Raczek, A.E., Ware, J.E., Bjorner, J.B., Gandek, B., Haley, S.M., Aaronson, N.K., ...Sullivan, M. (1998). Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA project. *Journal of Clinical Epidemiology*, 51, 1203-1214.
- Raji, M.A., Kuo, Y-F, Al Snih, S., Markides, K.S., Peek, M.K., & Ottenbacher, K.J. (2005). Cognitive status, muscle strength, and subsequent disability in older Mexican Americans. *Journal of the American Geriatrics Society*, 53, 1462–1468.
- Rantanen T, Era P, Kauppinen, M. & Heikkinen, E. (1994). Maximal isometric muscle strength and socio-economic status, health and physical activity in 75-year-old persons. *Journal of Aging and Physical Activity*, 2, 206-220.
- Rantanen ,T., Guralnik J.M., & Izmirlian, G. (1998).The association of muscle strength with maximal walking speed in disabled older women. *American Journal of Physical Medicine and Rehabilitation*, 77, 299-305.
- Rantanen, T, Guralnik, J. M., Foley, D., Masaki, K., Leveille, S., Curb, J. D., & White, L.(1999). Midlife hand grip strength as a predictor of old age disability, *Journal of the American Medical Association*, 281, 558-560.
- Rantanen, T., Guralnik, J. M., Ferrucci, L., Penninx, B. W. J. H., Leveille, S., Sipilä, S. & Fried, L. P. (2001), Coimpairments as Predictors of Severe Walking Disability in Older Women. *Journal of the American Geriatrics Society*, 49, 21–27. doi: 10.1046/j.1532-5415.2001.49005.
- Rantanen T, Avlund K, Suominen H, Schroll M, Frändin K, & Pertti E. (2002).Muscle strength as a predictor of onset of ADL dependence in people aged 75 years. *Aging Clinical and Experimental Research*, 14, 10-5.
- Rantanen, T., Volpato, S., Ferrucci, L., Heikkinen, E., Fried, L.P., & Guralnik, J.M. (2003). Handgrip strength and cause-specific and total mortality in older disabled women:exploring the mechanism. *Journal of the American Geriatric Society*, 51, 636-641.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Denmark, Copenhagen: Nielsen & Lydiche.
- Raven, J. C., Raven, J. E., & Court, J. H. (1977). *Progressive matrices*. Oxford, England: Oxford Psychologists Press.
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using MULTILOG. *Journal of Educational Measurement*, 27, 133-144.
- Reise, S. P., & Henson, J. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, 81, 93-103.
- Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*, 14, 95-101.
- Reuben, D.B, Rubenstein, L.V, Hirsch, S.H, Hays, R.D.(1992). Value of functional status as a predictor of mortality: results of a prospective study. *American Journal of Medicine*, 93, 663-9.
- Reuben, D. B., Seeman, T. E., Keeler, E., Hayes, R. P., Bowman, L., Sewall, A., Hirsch S. H., Wallace, R. B., & Guralnik, J. M. (2004). Refining the Categorization of physical functional status: the added value of combining self-reported and performance-based measures. *Journal of Gerontology series A Biological Science Medical Science*, 59, 1056-1061.
- Reuben, D.B, Valle L. A, Hays, R. D, & Sui, A.L. (1995). Measuring physical function in community-dwelling older persons: a comparison of self-administered, interview-administered, and performance-based measures. *Journal of the American Medical Association*, 273, 17-23.
- Reuben, D. B., Seeman, T. E., Hirsch, S. H., Keeler, E., Hayes, R. P., Bowman, L., Sewall, A., Wallace, R. B., & Guralnik, J. M. (2004). 'Refining the categorization of physical functional status: the added value of combining self-reported and performance-based measures'. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59, 1056-1061.
- Richardson, J., Mallen, C.D., & Burrell, H.S. (2008). A woman living with osteoarthritis: A case report. *Cases Journal*, 1, 153.
- Richardson, J., Chan, D., Risdon, K., Giles, C., Mulveney, S., & Cripps, D. (2008). Does monitoring change in function in community-dwelling older adults alter outcome? A randomized controlled trial. *Clinical Rehabilitation*, 22, 1061-1070.

- Rogers, W., & Miller, B. (1997). A comparative analysis of ADL questions in surveys of older people. *Journal of Gerontology, Series B: Psychological Science*, 52, 21-35.
- Roorda, L. D., Roebroek, M. E., van Tilburg, T., Lamkhorst, G. J., & Bouter, L. M. (2004). Measuring activity limitations in climbing stairs: development of a hierarchical scale for patients with lower-extremity disorders living at home. *Archives of Physical Medicine and Rehabilitation*, 85, 967-971.
- Roorda, L., Green, J., De Kluis, K., Molenaar, I., Bagley, P., Smith, J., & Guerts, A. (2008). Excellent cross-cultural validity, intra-test reliability and construct validity of the Dutch Rivermead Mobility Index in patients after stroke undergoing rehabilitation. *Journal of Rehabilitation Medicine*, 40, 727-732.
- Roorda, L., D. Scholtes, Van der Lee, J., Becher, J., & Dallmeijer, A., J., (2010). Measuring Mobility Limitations in Children With Cerebral Palsy: Development, Scalability, Unidimensionality, and Internal Consistency of the Mobility Questionnaire, MobQues47. *Archives of Physical Medicine and Rehabilitation*, 91, 1194-1209.
- Rosow, I., & Breslau, N. (1966). A Guttman health scale for the aged. *Journal of Gerontology*, 21, 556-559.
- Rydwik, E., Frandin, K., & Akner, G. (2004). Effects of physical training on physical performance in institutionalized elderly patients (70+) with multiple diagnosis. *Age and Ageing*, 33, 13-23.
- Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-Based Medicine : How to Practice and Teach EBM*. 2nd ed. London, England: Churchill Livingstone; 67-93.
- Salomonsson, B., Ahlstrom, S., Dalen, N., Lillkrona, U. (2009). The Western Ontario shoulder instability Index (WOSI): validity, reliability, and responsiveness retested with a Swedish translation. *Acta Orthopaedica*, 80, 233-238.
- Salthouse, T. A. (1992). What do adult age differences in the Digit Symbol Substitution test reflect? *Journal of Gerontology: Psychological Sciences*, 47, 121-128.
- Salthouse, T. A. (1993). Speed and knowledge as determinants of adult age differences in verbal tasks. *Journal of Gerontology*, 48, 29-36.
- Schiel, J. L., & Stocking, V. B. (2001). Benefits of TIP Summer Residential Program participation, as reflected by subsequent academic performance in high school. In N. Colangelo & S. G. Assouline (Eds.), *Talent Development IV: Proceedings of the 1998 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development* Scottsdale: Gifted Psychology Press.

- Schumacker, R. E. (2004). Rasch measurement using dichotomous scoring. *Applied Psychological Measurement*, 5, 328-349.
- Seeman, T. E., Berkman, L.F., Charpentier, P.A., Blazer, D.G., Albert, M.S. & Tinetti, M.E. (1995). Behavioral and Psychosocial Predictors of Physical Performance: MacArthur Studies of Successful Aging. *Journals of Gerontology A: Medical Sciences*, 50, 177-183.
- Sheehan, B., Philpot, M. & Banerjee, S. (2002). Attributions of physical symptoms in patients of an old age psychiatry service. *International Journal of Geriatric Psychiatry* 17, 61-64.
- Sheehan, T. J., DeChello, L. M., Garcia, R., Fifield, J., Rothfield, N., & Reisine, S. (2002). Measuring disability: application of the Rasch model to activities of daily living (ADL/IADL). *Journal of Outcome Measurement*, 5, 839-863.
- Shipley, B. A., Der, G., Taylor, M. D., Deary, I. J. (2006). Cognition and all-cause mortality across the entire adult age range: health and lifestyle survey. *Psychosomatic Medicine*, 68, 17-24.
- Shinkai, S., Watanabe, S., Kumagai, S., Fujiwara, Y., Amano, H., Yoshida, H., et al. (2000). Walking speed as a good predictor for the onset of functional dependence in a Japanese rural community population. *Age Ageing*, 29, 441-446.
- Shye, S. (1988). Inductive and deductive reasoning: A structural reanalysis of ability tests. *Journal of Applied Psychology*, 73, 308-311.
- Sidney, S. (2005) COPD and incident cardiovascular disease hospitalizations and mortality: Kaiser Permanente Medical Care Program. *Chest*, 128, 2068-2075.
- Sijtsma, K., Debets, P., & Molenaar, I.W. (1990). Mokken scale analysis for polychotomous items: Theory, a computer program and an empirical application. *Quality & Quantity*, 24, 173-188.
- Sijtsma, K. & Junker, B. W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology*, 49, 79-105.
- Sijtsma, K. & Hemker, B. T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika*, 63, 183-200.
- Sijtsma, K., & Molenaar, I. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks: Sage.

- Sinclair, A. J., Gray, L., Lunec, J., and Barnett, A. H. (1993). Serum Fluorescence “Protein Browning” in Elderly Patients with Diabetes Mellitus Age Ageing, 22, 15b-15.
- Sinclair, A. J., Conroy, S. P., Baye, r A. J., (2008) Impact of diabetes on physical function in older people. *Diabetes Care*, 31, 233-235.
- Sitzer, D.I., Twamley, E.W., & Jeste, D.V. (2006). Cognitive training in Alzheimer’s Disease: A meta-analysis of the literature. *Acta Psychiatrica Scandinavica*, 114, 75–90.
- Simonsick, E. M., Newman, A. B., Visser, M., Goodpaster, B., Kritchevsky, S. B., Rubin, S., et al. (2008). Mobility limitation in self-described well-functioning older adults: importance of endurance walk testing. *The Journals of Gerontology, Series A: A Biological Science*, 63, 841-47.
- Siu, A. L, Reuben, D. B, & Hays, R. D. (1990). Hierarchical measures of physical function in ambulatory geriatrics. *Journal of the American Geriatric Society*, 38, 1113-1119.
- Smith, E.V. Jr. (2005). Effect of item redundancy on Rasch item and person estimates. *Journal of Applied Measurement*, 6, 147-63.
- Skelton, D., Young, A. A. & Greig, A.C. (1997). Muscle function in women aged 65-89 years meeting two sets of health criteria. *Aging Clinical Experimental Research*. 9, 106-111.
- Sonn, U. (1996). Longitudinal studies of dependence in daily life activities among elderly persons. *Scandinavian Journal of Rehabilitation Supplement*, 34, 3-35.
- Song, J., Chang, R. W. (2006). "Population impact of arthritis on disability in older adults." *Arthritis and Rheumatism*, 55, 248-255.
- Spearman, C. (1904). “General intelligence” objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- Spector, W. D., & Fleishman, J. A. (1998). Combining activities of daily living with instrumental activities of daily living to measure functional disability. *Journal of Gerontology Series B: Psychological Science*, 53, 46-57.
- Staff, R.T., Murray, A. D., Deary, I.J. & Whalley, L.J. (2004). What provides cerebral reserve? *Brain*, 127, 1191-1199.

- Starkstein, S.E., Ingram, L., Garau, M. L., & Mizrahi, R. (2005). On the overlap between apathy and depression in dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76, 1070-1074.
- Startzell, J. K., Owens, D.A., Mulfinger, L.M., & Cavanagh, P.R. (2000), Stair negotiation in older people: a review. *Journal of the American Geriatric Society*, 48, 567-580.
- Stern, Y. (2003). The concept of cognitive reserve: a catalyst for research. *Journal of Clinical Experimental Neuropsychology*, 25, 589-93.
- Stern, Y. (2006). Cognitive Reserve and Alzheimer Disease. *Alzheimer Disability Association of Disorders*, 20, 69-74.
- Stern, Y.(2009). Cognitive reserve. *Neuropsychologia*, 47, 2015-28.
- Stucki, G., Daltroy, L., Katz, J.N., Johannesson, M., Liang, M.H. (1996). Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *Journal of Clinical Epidemiology*, 49,711-7.
- Stephens, R. (2006). Age-related decline in Digit-Symbol performance: Eye-movement and video analysis. *Archives of Clinical Neuropsychology*, 21, 101-107.
- Suri, P., Kiely, D. K., Leveille, S. G., Frontera, W. R., Bean, J. F. (2009) Trunk muscle attributes are associated with balance and mobility in older adults: a pilot study. *PM & R .Journal of Injury, Function, and Rehabilitation*, 1, 916-24.
- Taekema, D. G, Gussekloo, J, Maier, A. B, Westendorp, R. G, de Craen A. J. (2010).Handgripstrength as a predictor of functional, psychological and social health. Aprospective population-based study among the oldest old. *Age and Ageing*,39,331-7.
- Tao, W., Haley, S. M., Coster, W.J., Ni, P., & Jette, A.M. (2008).An exploratory analysis of functional staging using an item response theory approach. *Archives of Physical Medicine & Rehabilitation*, 89, 1046-1053.
- Teng, E. L., & Chui, H. C. (1987). The Modified Mini-Mental State (3MS) examination. *Journal of Clinical Psychiatry*, 48, 314-318.
- Tennant, A., Hillman, M., Fear, J., Pickering, A., & Chamberlain, M. A. (1996). Are we making the most of the Stanford Health Assessment Questionnaire? *British Journal of Rheumatology*, 35, 574-8.

- Teresi, J.A., Cross, P.S., & Golden, R.R. (1989). Some applications of latent trait analysis to the measurement of ADL. *Journal of Gerontology*, 44, 196-204.
- Tesfaye S, Chaturvedi N, Eaton SE, Ward JD, Manes C, Ionescu-Tirgoviste C.(2005). Vascular risk factors and diabetic neuropathy. *National English Journal of Medicine*, 352, 341-50.
- The Rasch model as a construct validation tool. [www.rasch.org/rmt/rmt221a.htm].
- Tinetti, M. E., Inouye, S. K., Gill, T. M., & Doucette, J. T. (1995). Shared risk factors for falls, incontinence, and functional dependence. Unifying the approach to geriatric syndromes. *Journal of the American Medical Association*, 273, 1348–1353.
- Tsutakawa, R. K. & Johnson, J. C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55, 371-390.
- Tucker, K. L., Bermudez, O. I., Gastaneda, G.(2000). Type 2 diabetes is prevalent and poorly controlled among Hispanic elders of Caribbean origin. *American Journal of Public Health*, 90, 1288.
- U.S. Department of Health and Human Services (1999). Older Adults and Mental Health. In: Mental Health: A Report of the Surgeon General. Available at: <http://www.surgeongeneral.gov/library/mentalhealth/chapter5/sec1.html> (accessed April 11, 2011).
- Valderrama-Gama, E., J. Damian, A. Ruigomerz, & Martin-Moreno, J.M. (2002). Chronic disease, functional status, and self-ascribed causes of disabilities among noninstitutionalized older people in Spain. *Journal of Gerontology. A Biological Science Medical Science* 57,16-21.
- Van Alphen, A., Halfens, R., Hasman A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than a good theory. *Journal of Advanced Nursing* , 20, 196–201.
- Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, 20, 1-19.
- Verbrugge, L. M, & Jette, A.M. (1994). The disablement process. *Social Science and Medicine*,38, 1-14.
- Verbrugge, L. M., Yang, L., & Juarez, L. (2004). Severity, timing, and structure of disability. *Soz.-Präventivmed*, 49, 110–121.
- Vittengl, J. R., White, C. N., McGovern, R.J., & Morton, B. J. (2006). Comparative validity of seven scoring systems for the instrumental activities of daily living scale in rural elders. *Aging and Mental Health*, 10, 40-47.

- Visser, M., Deeg, D.J., Lips, P., Harris, T., & Bouter, L. M. (2000). Skeletal muscle mass and muscle strength in relation to lower-extremity performance in older men and women. *Journal of American Geriatric Society*, 48, 381-386.
- Wang, L., van Belle, G., Kukull, W., & Larson, E.B. (2002). Predictors of functional change: A longitudinal study of nondemented people aged 65 and older. *Journal of the American Geriatrics Society*, 50, 1525-1534.
- Wang W.Z. (2002). Change of incidence of stroke after a community-based intervention for nine years in three cities in China. *Chinese Journal of Chronological Non-commune Disability*, 4, 30-33.
- Wang, M., Summers, J.A., Little, T., Turnbull, A., Poston, D. & Mannan, H. (2006). Perspectives of fathers and mothers of children with early intervention programmes in assessing family quality of life. *Journal of Intellectual Disability Research*, 50, 977-988.
- Wang CY, Sheu CF, & Protas E.(2007) Construct validity and physical performance of older adults in different hierarchical physical-disability levels. *Journal of Aging and Physical Activity*,15, 75-89.
- Ward, A. & Murray-Ward, M. (1994). Guidelines for the development of item banks. An NCME instructional module. *Educational Measurement: Issues and Practice*, 13, 34-39.
- Watson, R., Vander Ark, Lin. L. C., Fieo, R, Deary, I. J., & Meijer, R. R. Item response theory: how Mokken scaling can be used in clinical practice. *Journal of Clinical Nursing*, in press.
- Watson, R., Deary, I., Austin, E. (2007). Are personality trait items reliably more or less 'difficult'? Mokken scaling of the NEO-FFI. *Personality and Individual Differences*, 43, 1460–1469.
- Watson, R., & Deary, I. (2010). Reply to: A comment on Watson, Deary, and Austin (2007) and Watson, Roberts, Gow, and Deary (2008): How to investigate whether personality items form a hierarchical scale? *Personality and Individual Differences*, 48, 504–505
- Wechsler, D.M. (1981).Wechsler Adult Intelligence Scale-Revised (WAIS-R). San Antonio, TX: Psychological Corp.
- Weiner, M. F. (2006). Measuring QOL in dementia. *International Psychogeriatrics*, 18, 357-358.
- Weiss, J. E., and Ilowite, N. T. (2007). "Juvenile idiopathic arthritis". *Rheumatic Disability Clinical North America*, 33, 441–470.

- Wells, K. B., Hays, R. D., Burnam, M. A., Rogers, W., Greenfield S., & Ware, J. Jr. (1989). Detection of depressive disorder for patients receiving prepaid or fee-for-service care. Results from the Medical Outcomes Study. *Journal of the American Medical Association*, 262, 298–302.
- Whalley, L. J., Starr, J. M., Athawes R, D. Hunter, Pattie, A. & Deary, I. J. (2000). Childhood mental ability and dementia. *Neurology*, 55,1455–9.
- World Health Organization. *ICF: International Classification of Functioning, Disability and Health*. Geneva, Switzerland: World Health Organization; 2001.
- World Health Organization (2004) Disease incidence, prevalence and disability.WHO Statistical Informationwho.int/.../global...disease/GBD_report_2004update_part3.pdf
- Wieland, G., D. (2005) From Bedside to Bench: Research in Comorbidity and Aging. *Science Aging Knowledge Environment*. [DOI: 10.1126/sageke.39.pe29].
- Wu, S. Y., & Green, H. Y. (2000). Projection of chronic illness prevalence and cost inflation. Washington, DC: *RAND Health*.
- Wiener, J. M., & Tilly, J. (2002). Population ageing in the United States of America: implications for public programmes. *International Journal of Epidemiology*, 31, 776-81.
- Wild, S., Roglic, G., Green, A., Sicree, R., King, H. (2004). Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care*. 27,1047-1053.
- Williamson, J.D., Espeland, M., Kritchevsky, S.B., Newman, A.B., King, A.C., Pahor M., Guralnik, J.M., Pruitt, L.A., Miller, M.E., LIFE Study Investigators. (2009). Changes in cognitive function in a randomized trial of physical activity: results of the lifestyle interventions and independence for elders pilot study. *Journals of Gerontology Series A-Biological Sciences & Medical Sciences*, 64, 688-9.
- Wilms, H., Riedel-Heller, S. G., & Angermeyer, M. C. (2007). Limitations in activities of daily living and instrumental activities of daily living capacity in a representative sample: disentangling dementia-and mobility-related effects. *Comprehensive Psychiatry* , 48, 95-101.
- Wilson, M. (1989). comparison of deterministic and probabilistic approaches to measuring learning structures. *Australian Journal of Education*, 33,127-144.

- Wilson, M., Miller, D.K., Andresen, E.M., Malstrom, T.K., Miller, J.P., & Wolinsky, F.D.(2005). Fear of falling and related activity restriction among middle aged African Americans. *The Journals of Gerontology Series. A biological Sciences and Medical Sciences*, 60, 355-360.
- Wilson, M. (2005). *Constructing Measures: An Item Response Modelling Approach*. Mahwah, NJ: Erlbaum.
- Wilson, M. (2006). Improving measurement in health education and health behaviour research using item response modelling: comparison with the classical test theory approach. *Health Education Research*, 21, i19-32.
- Wirtz, M. A., & Voigt-Radloff, S. (2008). The international classification of functioning, disability and health-orientated occupational therapy assessment: a Rasch analysis of its domains. *International Journal of Rehabilitation Research*, 31, 23–32.
- Wolfe, F., & Kong, S. X. (1999). Rasch analysis of the Western Ontario MacMaster Questionnaire (WOMAC) in 2205 patients with osteoarthritis, rheumatoid arthritis, and fibromyalgia. *Annals of the Rheumatic Diseases*, 58, 563-568.
- White, L. R., Cartwright, W. S., Cornoni-Huntley, J., & Brock, D. B.(1986). Geriatric Epidemiology. *Annual Review of Gerontology and Geriatrics*, 6, 215-311.
- Wright, B. D., & Stone, M. H. (2004). *Making measures*. Chicago: Phaneron Press.
- Ylikoski, A., Erkinjuntti, T., Raininko, R., Sarna, S., Sulkava, R., & Tilvis, R. (1995). White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. *Stroke* 26, 1171–7.
- Young, A. (1999) Strength, Power, and Functional Performance in Old Age. In- *Proceedings of Muscle Power: Scientific and Practical Perspectives*. (Eds-S. Harridge, S. Dinan, R. Woledge). University of London. London UK.
- Zawacki, T., Grace, J., Paul, R., Moser, D., Ott, B., Gordan, N. & Cohen, R. (2002). Behaviroal problems as predictors of functional abilities of vascular dementia patients. *Journal of Neuropsychiatry and Clinical Neuroscience*, 14, 296-302.

Appendix

Table A1 Tasks of daily life adopted from the Health Interview Survey Supplement on Aging Questionnaire

Group 1: Mobility		Group 2: Complex		Group 3: Self-care		Group 4: Upper Extremity	
Task	(r)*	Task	(r)*	Task	(r)*	Task	(r)*
Walk 1/2 mile	0.77	Pay bills†	0.70	Using toilet†	0.72	Gripping	0.67
Climb 10 steps	0.73	Meal preparation†	0.69	Dressing†	0.62	Reaching	0.63
Transfer from bed to chair†	0.70	Shopping†	0.64	Bathing†	0.54		
Walk in own home†	0.63	Light housework†	0.54	Eating†	0.47		
Heavy housework†	0.54	Telephone use†	0.47				
Lifting and carrying 10 lb bag of groceries	0.51						
Proportion with difficulty in ≥ 1 task in group	36%		8%		3%		18%

Figure A1 Sample sheet of Townsend's Disability Scale

Townsend's Disability Scale	
Are you able to...	Score
1 Cut your own toe-nails?
2 Wash all over or bathe?
3 Get on a bus?
4 Go up and down stairs?
5 Do the heavy housework?
6 Shop and carry heavy bags?
7 Prepare and cook a hot meal?
8 Reach an overhead shelf?
9 Tie a good knot in a piece of string?
TOTAL
If answer is 'yes', ask 'Do you have any difficulty?'	
Score	
0 if without difficulty	
1 if with difficulty	
2 if unable	